```
     ||:::      :::   ||||||      :::                  :::
 237 rpGluTyrTyrSerArgGluLysProLysGlnGluGlyGluAlaLysVal 253

 205 CGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAGGGA 254
     ::::::|||     |||    :::                         ||
 254 GluAlaHisArgAlaGlyGlnGlyHisProCysArgAsnAlaGluAlaGl 270

 255 GGAAGGAGGGCCGAGG..............AGTGGAGGGGCTCAGGCGA 289
     ||||||||||||:::               :::|||:::|||      |
 270 uGluGlyGlyProGluThrThrPheValCysThrGlyAsnAlaPheLeuL 287

 290 AGCTGGGGTGCTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCC 339
     ||         :::|||   ||| |||             ||||||
 287 ys.........AlaTrpValTyr.Arg............ProGlyGluAs 296

 340 GACAGAAGATTCTGGACTCC......CCAGACGGGACCAGGAGAGGGACG 383
      ||||||:::::: |||:::         |||:::::::::::::   |||
 296 pThrGluGluGlu.AspAsnSerAspSerAspSerAlaGluGluAspThr 312

 384 GCATGAGCGACACACACAAACACAGAACCACACAGCCAGTCC........ 425
     |||         :::|||::::::   ||||||:::::::::
 313 Ala.........GlnThrGlyAlaThrProHisThrSerAlaPheLeuLy 326

 426 ................................CAGGAGCCCAGTAATG 441
                                     :::|||   |||::::
 326 sAlaTrpValTyrArgProGlyGluAspThrGluGluGluAspSerAspS 343

 442 GAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCT...... 485
     :::::      :::|||:::   |||      :::      |||
 343 erAspSerAlaGluGluAspThrAlaGlnThrGlyAlaThrProHisThr 359

 486 ........................,.ACACCTGGGCAGCAGACAGAAGA 508
                               ||||||:::::|||||||
 360 SerAlaPheLeuLysAlaTrpValTyrArgProGlyGluAspThrGluGl 376

 509 AGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGC 558
     |:::::::::                |||:::::::::      :
 376 uGluAsnSer.....................AspLeuAspSerAlaGluG 386

 559 AAGAGCTGCATCAGTCAAAC...ACCGGGGATAAATCTGGATTT...... 599
     :::::      |||::::::   |||         |||:::|||
 386 luAspThrAlaGlnThrGlyAlaThrProHisThrSerAlaPheLeuLys 402

 600 ...GGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTA 646
                |||   |||||||||   ||| :::||||||:::
 403 AlaTrpValTyrArgProGlyGluAsp...Thr.GluGluGluAsnSerA 418

 647 AAATGCCAGAAGCAGGTGAAGAGCAACCACAA 678
       :::   :::|||   |||:::    |||
 418 spLeuAspSerAlaGluGluAspThrAlaGln 428

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-821-818-3

seq_documentation_block:
; Sequence 3, Application US/08821818
; Patent No. [redacted]
;  GENERAL INFORMATION:
;    APPLICANT:  Fisher, Paul B.
;    TITLE OF INVENTION:  IDENTIFICATION OF THE PROGRESSION
;    TITLE OF INVENTION:  ELEVATED GENE-3 AND USES THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/821,818
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chan, Albert Wai-Kit
;      REGISTRATION NUMBER:  36,479
;      REFERENCE/DOCKET NUMBER:  51523
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  657 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-821-818-3


alignment_scores:
            Quality:   95.00              Length:    246
              Ratio:  0.805                Gaps:     15
 Percent Similarity:  47.967   Percent Identity: 24.797

alignment_block:
 US-09-651-563-808 x US-08-821-818-3   ..

 Align seq 1/1  to: US-08-821-818-3  from: 1  to: 657

      55 GCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAGTTG 104
         |||||||||   :::   :::      :::||||||||||:::::|||
     212 AlaThrGluGluLysGlyThrGluAsnLysAlaAspProSerAsnSerPr 228

     105 AAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTGTGT 154
          |||                          :::|||   :::      :::|
     228 oSer.......................SerGlySerHisSerArgAlaT 237

     155 GGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTCATG 204
         ||:::      :::   ||||||      :::                 :::
     237 rpGluTyrTyrSerArgGluLysProLysGlnGluGlyGluAlaLysVal 253

     205 CGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAGGGA 254
         ::::::|||      |||   :::                        ||
     254 GluAlaHisArgAlaGlyGlnGlyHisProCysArgAsnAlaGluAlaGl 270

     255 GGAAGGAGGGCCGAGG..............AGTGGAGGGGCTCAGGCGA 289
         ||||||||||||:::               :::|||:::|||      |
     270 uGluGlyGlyProGluThrThrPheValCysThrGlyAsnAlaPheLeuL 287

     290 AGCTGGGGTGCTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCC 339
         ||         :::|||   ||| |||             ||||||
     287 ys.........AlaTrpValTyr.Arg............ProGlyGluAs 296

     340 GACAGAAGATTCTGGACTCC......CCAGACGGGACCAGGAGAGGGACG 383
          ||||||:::::: |||:::         |||:::::::::::::   |||
     296 pThrGluGluGlu.AspAsnSerAspSerAspSerAlaGluGluAspThr 312

     384 GCATGAGCGACACACACAAACACAGAACCACACAGCCAGTCC........ 425
         |||         :::|||::::::   ||||||:::::::::
     313 Ala.........GlnThrGlyAlaThrProHisThrSerAlaPheLeuLy 326

     426 ................................CAGGAGCCCAGTAATG 441
                                         :::|||   |||::::
     326 sAlaTrpValTyrArgProGlyGluAspThrGluGluGluAspSerAspS 343

     442 GAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCT...... 485
         :::::      :::|||:::   |||      :::      |||
     343 erAspSerAlaGluGluAspThrAlaGlnThrGlyAlaThrProHisThr 359

     486 ........................ACACCTGGGCAGCAGACAGAAGA 508
                                   ||||||::::::|||||||||
```

```
     360 SerAlaPheLeuLysAlaTrpValTyrArgProGlyGluAspThrGluGl 376

     509 AGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGC 558
         |:::::::::                  |||::::::::::      :
     376 uGluAsnSer.....................AspLeuAspSerAlaGluG 386

     559 AAGAGCTGCATCAGTCAAAC...ACCGGGGATAAATCTGGATTT..... 599
         :::::      |||::::::   |||         |||:::|||
     386 luAspThrAlaGlnThrGlyAlaThrProHisThrSerAlaPheLeuLys 402

     600 ...GGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTA 646
                 |||   |||||||||   ||| :::|||||||:::
     403 AlaTrpValTyrArgProGlyGluAsp...Thr.GluGluGluAsnSerA 418

     647 AAATGCCAGAAGCAGGTGAAGAGCAACCACAA 678
         :::   :::|||   |||:::      |||
     418 spLeuAspSerAlaGluGluAspThrAlaGln 428

eq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-244-603A-1

eq_documentation_block:
 Sequence 1, Application US/08244603A
 Patent No. 6200768
  GENERAL INFORMATION:
    APPLICANT:  Mandelkow, Eva-Maria
    APPLICANT:  Mandelkow, Eckhard
    APPLICANT:  Lichtenberg-Kraag, Birgit
    APPLICANT:  Biernat, Jacek
    APPLICANT:  Drewes, Gerard
    APPLICANT:  Steiner, Barbara
    TITLE OF INVENTION:  No. 6200768el Tools For The Diagnosis And
    TITLE OF INVENTION:  Treatment Of Alzheimer's Disease
    NUMBER OF SEQUENCES:  1
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
      ADDRESSEE:  Borun
      STREET:  233 South Waker Drive, 6300 Sears Tower
      CITY:  Chicago
      STATE:  Illinois
      COUNTRY:  United States of America
      ZIP:  60606-6402
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Tape
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/244,603A
      FILING DATE:
      CLASSIFICATION:  435
    ATTORNEY/AGENT INFORMATION:
      NAME:  Joseph A. Williams, Jr.
      REGISTRATION NUMBER:  38,659
      REFERENCE/DOCKET NUMBER:  28384/32778
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  312-474-6300
      TELEFAX:  312-484-0448
      TELEX:  25-3856
  INFORMATION FOR SEQ ID NO:  1:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  441 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  protein
-08-244-603A-1


ignment_scores:
          Quality:   92.00                Length:    208
            Ratio:  0.829                  Gaps:     14
ercent Similarity:  53.365   Percent Identity: 28.365
```

```
alignment_block:
 US-09-651-563-808 x US-08-244-603A-1    ..

 Align seq 1/1  to: US-08-244-603A-1  from: 1  to: 441

      76 GGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGTGAGAGTGAAGA 125
         |||      |||:::  :::|||          ::::::   :::   |||::
     100 GlyThrThrAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAs 116

     126 GGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGTGACTCAGAGTGAGA 175
         :|||   ||||||                    ||||||||||::: |||
     116 pGluAlaAlaGly.................HisValThrGlnAla.Arg 126

     176 AGGCCCTCGAAGTCGTCGTCCCTCTCATGCGGTGCCACGCCCATGGACCT 225
               ||||||||||:::::            :::|||      |||
     127 MetValSerLysSerLysAsp............GlyThrGlySerAsp.. 138

     226 TCTTGTCTCGTCACGGCCATAACTAGGGAGGAAGGAGGGCCGAGGAGTGG 275
                                        |||:::::::::|||::::
     139 ..........................AspLysLysAlaLysGlyAlaA 146

     276 AGGGGCTCAGGCGA..............................AGC 292
         ::|||    :::                                 :::
     146 spGlyLysThrLysIleAlaThrProArgGlyAlaAlaProProGlyGln 162

     293 TGGGGTGCTGTTGGGGGTATCCGAGTC............CCAGAAGCACC 330
            |||   :::::::::   |||:::           |||   |||||
     163 LysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProProAlaPr 179

     331 TGGAACCCCGACAGAAGATTCTGGACTCCCCAGACGGGACCAGGAGAGGG 380
         |   ||||||      :::||||||   |||   ::::::   :::|||:
     179 oLysThrPro...ProSerSerGlyGluProProLysSerGlyAspArgS 195

     381 ACGGCATGAGCGACACACACAAACACA...GAACCACACAGCCAGTCCCA 427
         ::|||   ||| :::      :::         |||   |||:::|||::
     195 erGlyTyrSer.SerProGlySerProGlyThrProGlySerArgSerAr 211

     428 GGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCA......GCAGCTG 471
         :   ||||||         |||      :::||||||      :::|||
     211 gThrProSerLeuProThrProProThrArgGluProLysLysValAlaV 228

     472 AAAGTCGGGATCCTACACCTGGGCAGCAGACAGAAGAAGATCAG...GAT 518
            |||   |||        :::   ::::::   :::::::::   :::
     228 alValArgThrProProLysSerProSerSerAlaLysSerArgLeuGln 244

     519 ACAGCTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGCAAGAGCTGCA 568
         ||||||   :::|||:::   |||::: |||   :::   ||||||   |
     245 ThrAlaProValProMetProAspLeu.LysAsnVal...LysSerLysI 260

     569 TCAGTCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTG 618
         ||::::::|||   :::::::::   :::   |||:::|||   ||||||
     260 leGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly...LysVal 275

     619 AAGATAATACCTAAAGAG 636
         :::||||||   |||:::
     276 GlnIleIleAsnLysLys 281

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-893-852A-1

seq_documentation_block:
; Sequence 1, Application US/08893852A
; Patent No. 6080558
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  NEW HUMAN GROWTH REGULATOR PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
```

OM of: US-09-651-563-808 to: A_Geneseq_0601:* out_format : pfs

Date: Jun 28, 2001 11:18 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09651563/runat_28062001_145611_7551/app_query.fasta_1.854
-DB=A_Geneseq_0601 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09651563_@CGN1_1_51 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-651-563-808
Query length: 781
Database: A_Geneseq_0601:*
Database sequences: 412676
Database length: 60623988
Search time (sec): 73.750000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation	..
/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76877	+	887.00	1593.93	4.2e-81	160	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76869	+	550.00	986.19	3.8e-47	126	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76866	+	423.50	758.75	2.1e-34	108	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76870	+	423.50	758.75	2.1e-34	108	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB60500	+	233.00	413.49	3.4e-15	111	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83158	+	138.50	243.26	1.1e-05	102	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY52998	+	138.50	243.26	1.1e-05	102	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY12447	+	128.50	226.59	0.0001	89	!	H	
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83168	+	121.50	211.78	0.0006	109	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83169	+	120.00	212.45	0.0008	79	!	H	
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07749	+	109.00	186.06	0.0122	146	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB33006	+	103.50	165.59	0.0620	396	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW79958	+	103.00	160.69	0.0796	578	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY41104	+	103.00	160.69	0.0796	578	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY06514	+	103.00	160.69	0.0796	578	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY96255	+	102.50	152.43	0.1143	1162	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY58500	+	102.50	152.43	0.1143	1162	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB41231	-	102.50	142.53	0.1590	2971	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB50363	-	102.50	142.53	0.1590	2972	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB50362	-	102.50	142.03	0.1617	3118	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAW87503	-	100.50	148.36	0.1846	1212	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83167	+	100.00	175.22	0.0821	87	!	H	
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW57046	-	99.00	157.04	0.1787	411	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW79957	+	97.00	152.30	0.2952	457	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY41103	+	97.00	152.30	0.2952	457	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY06515	+	97.00	152.30	0.2952	457	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB42231	+	97.00	144.54	0.3826	955	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB26239	-	96.50	140.95	0.4700	1232	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB87791	-	96.00	166.10	0.2212	104	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY41102	+	96.00	146.67	0.4231	657	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW47601	+	95.50	163.95	0.2590	117	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW47602	+	95.50	163.95	0.2590	117	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW47603	+	95.50	163.95	0.2590	117	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83162	+	95.50	163.95	0.2590	117	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83163	+	95.50	163.95	0.2590	117	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83164	+	95.50	163.95	0.2590	117	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW47600	+	95.50	163.86	0.2598	118	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83161	+	95.50	163.86	0.2598	118	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW47599	+	95.00	163.13	0.2900	116	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB41949	+	94.50	148.92	0.5078	410	!		
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAW87504	-	94.50	138.90	0.7095	1061			
/SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:AAR08423	-	93.50	155.27	0.4878	189			
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG07240	+	93.00	147.58	0.6873	360			
/SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:AAR47811	+	92.50	142.96	0.8738	512			
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83159	+	92.00	157.70	0.5821	116			

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76877

seq_documentation_block:
ID AAB76877 standard; Protein; 160 AA.
XX
AC AAB76877;
XX
DT 12-APR-2001 (first entry)
XX
DE Human lung tumour protein related protein sequence SEQ ID NO:809.
XX
KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW cytostatic; antisense inhibition.
XX
OS Homo sapiens.
XX
PN WO200100828-A2.
XX
PD 04-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18061.
XX
PR 30-JUN-1999; 99US-0346492.
PR 15-OCT-1999; 99US-0419356.
PR 17-DEC-1999; 99US-0466867.
PR 30-DEC-1999; 99US-0476300.
PR 06-MAR-2000; 2000US-0519642.
PR 22-MAR-2000; 2000US-0533077.
PR 10-APR-2000; 2000US-0546259.
PR 27-APR-2000; 2000US-0560406.
PR 05-JUN-2000; 2000US-0589184.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
PI Retter MW, Mannion J;
XX
DR WPI; 2001-071488/08.
XX
PT Lung tumor-associated proteins and the nucleic acids that encode them,
PT useful for preventing, diagnosing and treating lung cancer -
XX
PS Example 1; Page 427; 436pp; English.
XX
CC The present invention describes immunogenic portions of lung tumour-
CC associated proteins (I) and the nucleic acids (NAs) that encode them.
CC (I) have cytostatic activity and can be used in gene therapy, antisense
CC inhibition and in vaccines. The NAs and the lung tumour-associated
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with their inappropriate expression,
CC especially lung cancers. For example, the NAs may be administered to
CC treat diseases by rectifying mutations or deletions in a patient's genome
CC that affect the activity of the protein by expressing inactive proteins
CC or to supplement the patients own production of (I). Additionally, the
CC NAs may be used to produce the lung-tumour associated protein, according
CC to standard recombinant DNA methodology. Conversely, antisense NA
CC molecules may be administered to down regulate protein expression by
CC binding with the cells own genes and preventing their expression. The NA
CC and complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar NA sequences in
CC samples, and hence which patients may be in need of treatment for lung
CC cancer. The (I) may be used as antigens in the production of antibodies
CC and in assays to identify modulators (agonists and antagonists) of the
CC expression and activity of the protein. AAF68083 to AAF68878 and
CC AAB76848 to AAB76878 represent human lung tumour protein related
CC nucleotide and protein sequences which are used in the exemplification

```
CC   of the present invention.
XX
SQ   Sequence   160 AA;


alignment_scores:
            Quality:  887.00              Length:    160
              Ratio:  5.544                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-651-563-808 x AAB76877   ..

 Align seg 1/1  to: AAB76877  from: 1  to: 160

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 gGluGluGlyGlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysC 34

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysTrpGlyTyrProSerProArgSerThrTrpAsnProAspArgArgPhe 50

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TrpThrProGlnThrGlyProGlyGluGlyArgHisGluArgHisThrGl 67

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 84

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 117

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 150

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||||
     151 ProGluAlaGlyGluGluGlnProGlnVal 160

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76869

seq_documentation_block:
ID   AAB76869 standard; Protein; 126 AA.
XX
AC   AAB76869;
XX
DT   12-APR-2001  (first entry)
XX
DE   Human lung tumour protein related protein sequence SEQ ID NO:791.
XX
KW   Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW   lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW   cytostatic; antisense inhibition.
XX
OS   Homo sapiens.
XX
PN   WO200100828-A2.
XX
PD   04-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18061.
XX
PR   30-JUN-1999;   99US-0346492.
PR   15-OCT-1999;   99US-0419356.
PR   17-DEC-1999;   99US-0466867.
PR   30-DEC-1999;   99US-0476300.
PR   06-MAR-2000; 2000US-0519642.
PR   22-MAR-2000; 2000US-0533077.
PR   10-APR-2000; 2000US-0546259.
PR   27-APR-2000; 2000US-0560406.
PR   05-JUN-2000; 2000US-0589184.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Bangur CS,  Lodes MJ,  Fanger GR,  Vedvick TS,  Carter D;
PI   Retter MW,  Mannion J;
XX
DR   WPI; 2001-071488/08.
XX
PT   Lung tumor-associated proteins and the nucleic acids that encode them,
PT   useful for preventing, diagnosing and treating lung cancer -
XX
PS   Example 1; Page 413; 436pp; English.
XX
CC   The present invention describes immunogenic portions of lung tumour-
CC   associated proteins (I) and the nucleic acids (NAs) that encode them.
CC   (I) have cytostatic activity and can be used in gene therapy, antisense
CC   inhibition and in vaccines. The NAs and the lung tumour-associated
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with their inappropriate expression,
CC   especially lung cancers. For example, the NAs may be administered to
CC   treat diseases by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of the protein by expressing inactive proteins
CC   or to supplement the patients own production of (I). Additionally, the
CC   NAs may be used to produce the lung-tumour associated protein, according
CC   to standard recombinant DNA methodology. Conversely, antisense NA
CC   molecules may be administered to down regulate protein expression by
CC   binding with the cells own genes and preventing their expression. The NA
CC   and complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar NA sequences in
CC   samples, and hence which patients may be in need of treatment for lung
CC   cancer. The (I) may be used as antigens in the production of antibodies
CC   and in assays to identify modulators (agonists and antagonists) of the
CC   expression and activity of the protein. AAF68083 to AAF68878 and
CC   AAB76848 to AAB76878 represent human lung tumour protein related
CC   nucleotide and protein sequences which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   126 AA;


alignment_scores:
            Quality:  550.00              Length:    227
              Ratio:  4.583                 Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x AAB76869   ..

 Align seg 1/1  to: AAB76869  from: 1  to: 126

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         ||||||||||||||||||
      23 rProLeuLeuArgHisGly............................... 29

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151
```

```
 29 ..................................................  29

152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

 29 ..................................................  29

202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

 29 ..................................................  29

252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

 29 ..................................................  29

302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

 29 ..................................................  29

352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                          :::  ||||||||
 30 .................................Gly..HisThrGl  33

402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL  50

452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg  66

502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa  83

552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
     ||||||||||||||||||||||||||||||||||||||||||||||||||
100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
     |||||||||||||||||||||||||||||
117 ProGluAlaGlyGluGluGlnProGlnVal 126
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76866

seq_documentation_block:

```
ID   AAB76866 standard; Protein; 108 AA.
XX
AC   AAB76866;
XX
DT   12-APR-2001  (first entry)
XX
DE   Human lung tumour protein related protein sequence SEQ ID NO:786.
XX
KW   Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW   lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW   cytostatic; antisense inhibition.
XX
OS   Homo sapiens.
XX
PN   WO200100828-A2.
XX
PD   04-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18061.
XX
PR   30-JUN-1999;   99US-0346492.
PR   15-OCT-1999;   99US-0419356.
PR   17-DEC-1999;   99US-0466867.
PR   30-DEC-1999;   99US-0476300.
PR   06-MAR-2000; 2000US-0519642.
PR   22-MAR-2000; 2000US-0533077.
PR   10-APR-2000; 2000US-0546259.
PR   27-APR-2000; 2000US-0560406.
PR   05-JUN-2000; 2000US-0589184.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Bangur CS,  Lodes MJ,  Fanger GR,  Vedvick TS,  Carter D;
PI   Retter MW,  Mannion J;
XX
DR   WPI; 2001-071488/08.
XX
PT   Lung tumor-associated proteins and the nucleic acids that encode them,
PT   useful for preventing, diagnosing and treating lung cancer -
XX
PS   Example 1; Page 410; 436pp; English.
XX
CC   The present invention describes immunogenic portions of lung tumour-
CC   associated proteins (I) and the nucleic acids (NAs) that encode them.
CC   (I) have cytostatic activity and can be used in gene therapy, antisense
CC   inhibition and in vaccines. The NAs and the lung tumour-associated
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with their inappropriate expression,
CC   especially lung cancers. For example, the NAs may be administered to
CC   treat diseases by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of the protein by expressing inactive proteins
CC   or to supplement the patients own production of (I). Additionally, the
CC   NAs may be used to produce the lung-tumour associated protein, according
CC   to standard recombinant DNA methodology. Conversely, antisense NA
CC   molecules may be administered to down regulate protein expression by
CC   binding with the cells own genes and preventing their expression. The NA
CC   and complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar NA sequences in
CC   samples, and hence which patients may be in need of treatment for lung
CC   cancer. The (I) may be used as antigens in the production of antibodies
CC   and in assays to identify modulators (agonists and antagonists) of the
CC   expression and activity of the protein. AAF68083 to AAF68878 and
CC   AAB76848 to AAB76878 represent human lung tumour protein related
CC   nucleotide and protein sequences which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   108 AA;
```

alignment_scores:

```
           Quality:  423.50              Length:    215
             Ratio:  4.152                 Gaps:      3
 Percent Similarity:  47.442   Percent Identity: 46.977
```

alignment_block:

US-09-651-563-808 x AAB76866 ..

Align seg 1/1 to: AAB76866 from: 1 to: 108

```
  2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
    |||||||||||||||||||||||||||||||||||||||||||||||||||
  1 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 17

 52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
    |||||||||||||||||
 17 rProLeuLeuArgHisGly.............................. 23

102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

 23 ..................................................  23

152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

 23 ..................................................  23

202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
```

```
      23 ..................................................  23

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      23 ..................................................  23

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      23 ..................................................  23

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                          :::  ||||||||
      24 ....................................Gly..HisThrGl 27

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      27 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 44

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      44 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 60

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTG.............. 536
         ||||||||||||||||||||||||||||||||
      61 GlnLysLysIleArgIleGlnLeuArgSerGlnValLeuGlyArgGluMe 77

     537 .CGCGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGG 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      77 tArgAspMetGluGlyAspLeuGlnGluLeuHisGlnSerAsnThrGlyA 94

     586 ATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACC 629
         |||||||||||||||||||||||||||||||||||||||||||
      94 spLysSerGlyPheGlyPheArgArgGlnGlyGluAspAsnThr 108

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76870

seq_documentation_block:
ID   AAB76870 standard; Protein; 108 AA.
XX
AC   AAB76870;
XX
DT   12-APR-2001  (first entry)
XX
DE   Human lung tumour protein related protein sequence SEQ ID NO:793.
XX
KW   Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW   lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW   cytostatic; antisense inhibition.
XX
OS   Homo sapiens.
XX
PN   WO200100828-A2.
XX
PD   04-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18061.
XX
PR   30-JUN-1999;   99US-0346492.
PR   15-OCT-1999;   99US-0419356.
PR   17-DEC-1999;   99US-0466867.
PR   30-DEC-1999;   99US-0476300.
PR   06-MAR-2000; 2000US-0519642.
PR   22-MAR-2000; 2000US-0533077.
PR   10-APR-2000; 2000US-0546259.
PR   27-APR-2000; 2000US-0560406.
PR   05-JUN-2000; 2000US-0589184.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Bangur CS,  Lodes MJ,  Fanger GR,  Vedvick TS,  Carter D;
PI   Retter MW,  Mannion J;
XX
DR   WPI; 2001-071488/08.
XX
PT   Lung tumor-associated proteins and the nucleic acids that encode them,
PT   useful for preventing, diagnosing and treating lung cancer -
XX
PS   Example 1; Page 413-414; 436pp; English.
XX
CC   The present invention describes immunogenic portions of lung tumour-
CC   associated proteins (I) and the nucleic acids (NAs) that encode them.
CC   (I) have cytostatic activity and can be used in gene therapy, antisense
CC   inhibition and in vaccines. The NAs and the lung tumour-associated
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with their inappropriate expression,
CC   especially lung cancers. For example, the NAs may be administered to
CC   treat diseases by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of the protein by expressing inactive proteins
CC   or to supplement the patients own production of (I). Additionally, the
CC   NAs may be used to produce the lung-tumour associated protein, according
CC   to standard recombinant DNA methodology. Conversely, antisense NA
CC   molecules may be administered to down regulate protein expression by
CC   binding with the cells own genes and preventing their expression. The NA
CC   and complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar NA sequences in
CC   samples, and hence which patients may be in need of treatment for lung
CC   cancer. The (I) may be used as antigens in the production of antibodies
CC   and in assays to identify modulators (agonists and antagonists) of the
CC   expression and activity of the protein. AAF68083 to AAF68878 and
CC   AAB76848 to AAB76878 represent human lung tumour protein related
CC   nucleotide and protein sequences which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   108 AA;


alignment_scores:
           Quality:  423.50              Length:    215
             Ratio:  4.152                Gaps:      3
 Percent Similarity:  47.442   Percent Identity: 46.977

alignment_block:
 US-09-651-563-808 x AAB76870   ..

 Align seg 1/1  to: AAB76870  from: 1  to: 108

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 17

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         |||||||||||||||||
      17 rProLeuLeuArgHisGly.............................. 23

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

      23 ..................................................  23

     152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

      23 ..................................................  23

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

      23 ..................................................  23

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      23 ..................................................  23

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      23 ..................................................  23

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
```

```
                                        :::  ||||||||
     24 .....................................Gly..HisThrGl 27

    402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     27 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 44

    452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     44 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 60

    502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTG..............  536
        ||||||||||||||||||||||||||||||||||
     61 GlnLysLysIleArgIleGlnLeuArgSerGlnValLeuGlyArgGluMe 77

    537 .CGCGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGG 585
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     77 tArgAspMetGluGlyAspLeuGlnGluLeuHisGlnSerAsnThrGlyA 94

    586 ATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACC 629
        ||||||||||||||||||||||||||||||||||||||||||
     94 spLysSerGlyPheGlyPheArgArgGlnGlyGluAspAsnThr 108
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB60500

seq_documentation_block:
ID AAB60500 standard; Protein; 111 AA.
XX
AC AAB60500;
XX
DT 24-APR-2001 (first entry)
XX
DE Human cell cycle and proliferation protein CCYPR-48. SEQ ID NO:48.
XX
KW Cell cycle and proliferation protein; CCYPR; human; agonist;
KW antagonist; gene therapy; detection; gene therapy;
KW transgenic animal disease model; immune disorder;
KW developmental disorder; cell signalling disorder;
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KW arteriosclerosis; asthma; allergy; diabetes mellitus;
KW menstrual cycle disorder; bacterial infection.
XX
OS Homo sapiens.
XX
PN WO200107471-A2.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000WO-US19948.
XX
PR 21-JUL-1999; 99US-0145075.
PR 08-SEP-1999; 99US-0153129.
PR 10-NOV-1999; 99US-0164647.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
XX
DR WPI; 2001-112727/12.
DR N-PSDB; AAF59637.
XX
PT Human cell cycle and proliferation proteins and polynucleotides are
PT used to treat, diagnose and prevent immune, developmental and cell
PT signaling disorders and cell proliferative disorders including cancer -
XX
PS Claim 1; Page 158; 205pp; English.
XX
CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC associated with decreased expression of functional CCYPR, while CCYPR
CC antagonists are used to treat diseases or conditions associated with
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoasays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC Diseases which can be diagnosed, treated and prevented using CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
CC caused by bacteria.
XX
SQ Sequence 111 AA;

alignment_scores:
Quality: 233.00 Length: 85
Ratio: 3.426 Gaps: 1
Percent Similarity: 80.000 Percent Identity: 56.471

alignment_block:
US-09-651-563-808 x AAB60500 ..

Align seg 1/1 to: AAB60500 from: 1 to: 111

```
    429 GAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCG 478
        ||||||::::::   ||||||:::::::|||:::|||   ::::::|||||
     28 GluProThrAspGluGluProLysGluGluLysProProThrLysSerAr 44

    479 GGATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGA 528
        |:::|||||||||   |||:::   |||:::||||||   :::||||||||
     44 gAsnProThrProAspGlnLysArgGluAspAspGlnGlyAlaAlaGluI 61

    529 TCCCAGTGCGCGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAAC 578
        ||   |||   |||:::|||:::|||||||||||||||   |||::::::
     61 leGlnValProAspLeuGluAlaAspLeuGlnGluLeuCysGlnThrLys 77

    579 ACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATAC 628
        |||||||||  :::   :::   |||:::   ||||||   ||||||:::|
     78 ThrGlyAsp.GlyCysGluGlyGlyThrAspValLysGlyLysIleLeuP 94

    629 CTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCACAA 678
        |||||   ||||||   ||||||||||||||||||||   :::   |||
     94 roLysAlaGluHisPheLysMetProGluAlaGlyGluGlyLysSerGln 110

    679 GTT 681
        |||
    111 Val 111
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83158

seq_documentation_block:
ID AAY83158 standard; Protein; 102 AA.
XX
AC AAY83158;
XX
DT 24-JUL-2000 (first entry)
XX
DE PAGE-4 polypeptide.
XX
KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW immunoconjugate.
XX
OS Homo sapiens.
XX

PN WO200012706-A1.
XX
PD 09-MAR-2000.
XX
PF 31-AUG-1999; 99WO-US20046.
XX
PR 01-SEP-1998; 98US-0098993.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pastan I, Brinkmann U, Vasmatzis G, Lee B;
XX
DR WPI; 2000-237869/20.
DR N-PSDB; AAZ93524.
XX
PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT lymphocyte response and for raising antibodies which can be used to
PT detect the presence of PAGE-4 in cell samples or body tissues
XX
PS Disclosure; Figure 1; 63pp; English.
XX
CC PAGE-4 is a gene preferentially expressed in normal male and female
CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC and placenta, as well as in prostate cancer, testicular cancer and
CC uterine cancer. This expression pattern makes it a target for
CC diagnosis and for vaccine based therapy of such neoplasms.
CC An isolated PAGE-4 peptide which induces a cytotoxic T
CC lymphocyte response when bound to a major histocompatibility complex
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC immunogenic compositions to raise a cytotoxic T lymphocyte response
CC against cells expressing PAGE-4 including cancer cells of the
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC PAGE-4 peptide fragments can also be used in these compositions.
CC Antibodies against PAGE-4 and its peptide fragments can be used in
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC samples or body tissues. The presence of PAGE-4 in tissues which are
CC not related to reproduction can be indicative of the spread of
CC cancerous reproductive tissue. PAGE-4 can also be used to raise
CC antibodies which are then used as the targeting group of
CC immunoconjugates comprising toxins used in therapeutic applications.
CC This has applications for drug delivery systems.
XX
SQ Sequence 102 AA;

alignment_scores:
Quality: 138.50 Length: 79
Ratio: 2.770 Gaps: 2
Percent Similarity: 63.291 Percent Identity: 43.038

alignment_block:
US-09-651-563-808 x AAY83158 ..

Align seg 1/1 to: AAY83158 from: 1 to: 102

```
441 GGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACC 490
    ||||||   |||:::|||||||||   :::::::::::|||      ||
 28 GlyGluSerGlnGlnGluGluProProThrAspAsnGlnAspIleGluPr 44

491 TGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCG 540
    |||||||::   |||           |||      |||      |||
 44 oGlyGlnGluArgGlu.........GlyThrProProIleGluGluArgL 58

541 ACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAA 590
      :::|||||||||   ||||||:::      :::::|||   ::::::
 58 ysValGluGlyAspCysGlnGluMetAspLeuGluLysThrArgSerGlu 74

591 TCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAAC 640
       |||    ||||||   ||||||   |||      |||:::   :::|
 75 ArgGlyAsp.GlySerAspValLysGluLysThrProProAsnProLysH 91

641 ACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCA 675
    ||:::|||      ||||||||:::   ||||||
```

91 isAlaLysThrLysGluAlaGlyAspGlyGlnPro 102

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY52998

seq_documentation_block:
ID AAY52998 standard; Protein; 102 AA.
XX
AC AAY52998;
XX
DT 29-FEB-2000 (first entry)
XX
DE Human secreted protein clone bn365_53 protein sequence SEQ ID NO:2.
XX
KW Human; secreted protein; nutritional; cytokine; cell proliferation;
KW differentiation; immune stimulating; vaccine; suppression;
KW haematopoiesis regulation; tissue growth; activin; inhibin;
KW chemotactic; chemokinetic; haemostatic; thrombolytic; receptor;
KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW tumour inhibition; gene therapy.
XX
OS Homo sapiens.
XX
PN WO9957132-A1.
XX
PD 11-NOV-1999.
XX
PF 07-MAY-1999; 99WO-US09970.
XX
PR 07-MAY-1998; 98US-0084564.
PR 02-JUN-1998; 98US-0087645.
PR 22-JUL-1998; 98US-0093712.
PR 31-JUL-1998; 98US-0094935.
PR 10-AUG-1998; 98US-0095880.
PR 11-AUG-1998; 98US-0096068.
PR 06-MAY-1999; 99US-0096068.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;
PI DiBlasio-Smith E, Widom A;
XX
DR WPI; 2000-052937/04.
DR N-PSDB; AAZ33316.
XX
PT New polynucleotides encoding secreted human proteins, derived from
PT adult placenta, adult retina, fetal brain, fetal -
XX
PS Claim 9; Page 360-361; 492pp; English.
XX
CC The present invention describes new human secreted proteins which were
CC isolated from adult placenta, adult retina, foetal brain, foetal kidney,
CC adult blood, adult brain, adult thyroid, adult bladder, adult neural
CC tissue, adult testes, and adult lymph node cDNA libraries. The human
CC secreted proteins, and the polynucleotides encoding them, are predicted
CC to have biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals. Suggested activities include nutritional activity, cytokine
CC and cell proliferation/differentiation activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, and tumour inhibition
CC activity. The polynucleotides are also stated to be useful for gene
CC therapy. AAZ33316 to AAZ33373 encode human secreted proteins, and
CC AAY52998 to AAY53060 represent human secreted proteins, given in the
CC present invention.
XX
SQ Sequence 102 AA;

alignment_scores:

```
           Quality:  138.50            Length:     79
             Ratio:  2.770               Gaps:      2
 Percent Similarity:  63.291   Percent Identity: 43.038

alignment_block:
 US-09-651-563-808 x AAY52998   ..

 Align seq 1/1  to: AAY52998  from: 1  to: 102

     441 GGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACC 490
         ||||||   |||:::|||||||||   :::::::::::|||     ||
      28 GlyGluSerGlnGlnGluGluProProThrAspAsnGlnAspIleGluPr 44

     491 TGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCG 540
         |||||||:::   |||           |||      |||      |||
      44 oGlyGlnGluArgGlu.........GlyThrProProIleGluGluArgL 58

     541 ACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAA 590
           :::|||||||||   |||||||:::      :::::::|||   ::::::
      58 ysValGluGlyAspCysGlnGluMetAspLeuGluLysThrArgSerGlu 74

     591 TCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAAC 640
         |||    ||||||   ||||||  |||      |||:::   :::|
      75 ArgGlyAsp.GlySerAspValLysGluLysThrProProAsnProLysH 91

     641 ACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCA 675
         ||:::|||      ||||||||||:::   ||||||
      91 isAlaLysThrLysGluAlaGlyAspGlyGlnPro 102

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY12447

seq_documentation_block:
ID   AAY12447 standard; Protein; 89 AA.
XX
AC   AAY12447;
XX
DT   17-JUN-1999  (first entry)
XX
DE   Human 5' EST secreted protein SEQ ID NO:478.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS   Homo sapiens.
XX
PN   WO9906548-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01222.
XX
PR   01-AUG-1997;   97US-0905135.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153778/13.
DR   N-PSDB; AAX41280.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
XX
PS   Claim 27; Page 778; 824pp; English.
XX
CC   AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12261 to
CC   AAY12514, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence   89 AA;


alignment_scores:
           Quality:  128.50            Length:     76
             Ratio:  2.677               Gaps:      2
 Percent Similarity:  63.158   Percent Identity: 42.105

alignment_block:
 US-09-651-563-808 x AAY12447   ..

 Align seq 1/1  to: AAY12447  from: 1  to: 89

     441 GGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACC 490
         ||||||   |||:::|||||||||   :::::::::::|||     ||
      15 GlyGluSerGlnGlnGluGluProProThrAspAsnGlnAspIleGluPr 31

     491 TGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCG 540
         |||||||:::   |||           |||      |||      |||
      31 oGlyGlnGluArgGlu.........GlyThrProProIleGluGluArgL 45

     541 ACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAA 590
           :::|||||||||   |||||||:::      :::::::|||   ::::::
      45 ysValGluGlyAspCysGlnGluMetAspLeuGluLysThrArgSerGlu 61

     591 TCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAAC 640
         |||    ||||||   ||||||  |||      |||:::   :::|
      62 ArgGlyAsp.GlySerAspValLysGluLysThrProProAsn***LysH 78

     641 ACTGTAAAATGCCAGAAGCAGGTGAA 666
         ||:::|||      |||||||||:::
      78 isAlaLysThrLysGluAlaGlyAsp 86

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83168

seq_documentation_block:
ID   AAY83168 standard; Protein; 109 AA.
XX
AC   AAY83168;
XX
DT   24-JUL-2000  (first entry)
XX
DE   PAGE2 polypeptide.
XX
KW   PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW   uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW   cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW   immunoconjugate.
XX
OS   Homo sapiens.
XX
PN   WO200012706-A1.
XX
PD   09-MAR-2000.
XX
PF   31-AUG-1999;   99WO-US20046.
XX
PR   01-SEP-1998;   98US-0098993.
```

```
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan I,  Brinkmann U,  Vasmatzis G,  Lee B;
XX
DR   WPI; 2000-237869/20.
XX
PT   Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT   lymphocyte response and for raising antibodies which can be used to
PT   detect the presence of PAGE-4 in cell samples or body tissues
XX
PS   Disclosure; Figure 1b; 63pp; English.
XX
CC   PAGE-4 is a gene preferentially expressed in normal male and female
CC   reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC   and placenta, as well as in prostate cancer, testicular cancer and
CC   uterine cancer. This expression pattern makes it a target for
CC   diagnosis and for vaccine based therapy of such neoplasms.
CC   An isolated PAGE-4 peptide which induces a cytotoxic T
CC   lymphocyte response when bound to a major histocompatibility complex
CC   (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC   immunogenic compositions to raise a cytotoxic T lymphocyte response
CC   against cells expressing PAGE-4 including cancer cells of the
CC   prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC   PAGE-4 peptide fragments can also be used in these compositions.
CC   Antibodies against PAGE-4 and its peptide fragments can be used in
CC   detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC   samples or body tissues. The presence of PAGE-4 in tissues which are
CC   not related to reproduction can be indicative of the spread of
CC   cancerous reproductive tissue. PAGE-4 can also be used to raise
CC   antibodies which are then used as the targeting group of
CC   immunoconjugates comprising toxins used in therapeutic applications.
CC   This has applications for drug delivery systems. This sequence is
CC   of the PAGE2 polypeptide which shares sequence similarity with
CC   PAGE-4.
XX
SQ   Sequence   109 AA;


alignment_scores:
            Quality:  121.50              Length:     84
              Ratio:  2.430                Gaps:      2
 Percent Similarity:  59.524   Percent Identity: 38.095

alignment_block:
 US-09-651-563-808 x AAY83168   ..

 Align seg 1/1  to: AAY83168  from: 1  to: 109

     426 CAGGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAG 475
         |||||||||::::::   :::   |||:::|||||||||   ::::::::
      30 GlnGluProThrGluGluLysArgGlnGlnGluGluProProThrAspAs 46

     476 TCGGGATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGATACAGCTG 525
         :::::|||      ||||||||||:::   |||          |||
      46 nGlnAspIleGluProGlyGlnGluArgGlu.........GlyThrProP 60

     526 AGATCCCAGTGCGCGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCA 575
           |||      |||   :::|||||||||   ||||||:::
      60 roIleGluGluArgLysValGluGlyAspCysGlnGluMetAlaLeuLeu 76

     576 AACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAA 625
         :::      |||:::   |||    |||      |||:::      |||:
      77 LysIleGluAspGluProGlyAsp.GlyProAspValArgGluGlyIleM 93

     626 TACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCA 675
         ::|||      :::     |||:::   |||||||||:::   ||||||
      93 etProThrPheAspLeuThrLysValLeuGluAlaGlyAspAlaGlnPro 109

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY83169

seq_documentation_block:
ID   AAY83169 standard; Protein; 79 AA.

XX
AC   AAY83169;
XX
DT   24-JUL-2000  (first entry)
XX
DE   PAGE3 polypeptide.
XX
KW   PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW   uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW   cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW   immunoconjugate.
XX
OS   Homo sapiens.
XX
PN   WO200012706-A1.
XX
PD   09-MAR-2000.
XX
PF   31-AUG-1999;   99WO-US20046.
XX
PR   01-SEP-1998;   98US-0098993.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan I,  Brinkmann U,  Vasmatzis G,  Lee B;
XX
DR   WPI; 2000-237869/20.
XX
PT   Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT   lymphocyte response and for raising antibodies which can be used to
PT   detect the presence of PAGE-4 in cell samples or body tissues
XX
PS   Disclosure; Figure 1b; 63pp; English.
XX
CC   PAGE-4 is a gene preferentially expressed in normal male and female
CC   reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC   and placenta, as well as in prostate cancer, testicular cancer and
CC   uterine cancer. This expression pattern makes it a target for
CC   diagnosis and for vaccine based therapy of such neoplasms.
CC   An isolated PAGE-4 peptide which induces a cytotoxic T
CC   lymphocyte response when bound to a major histocompatibility complex
CC   (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC   immunogenic compositions to raise a cytotoxic T lymphocyte response
CC   against cells expressing PAGE-4 including cancer cells of the
CC   prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC   PAGE-4 peptide fragments can also be used in these compositions.
CC   Antibodies against PAGE-4 and its peptide fragments can be used in
CC   detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC   samples or body tissues. The presence of PAGE-4 in tissues which are
CC   not related to reproduction can be indicative of the spread of
CC   cancerous reproductive tissue. PAGE-4 can also be used to raise
CC   antibodies which are then used as the targeting group of
CC   immunoconjugates comprising toxins used in therapeutic applications.
CC   This has applications for drug delivery systems. This sequence is
CC   of the PAGE3 polypeptide which shares sequence similarity with
CC   PAGE-4.
XX
SQ   Sequence   79 AA;


alignment_scores:
            Quality:  120.00              Length:     77
              Ratio:  2.353                Gaps:      2
 Percent Similarity:  66.234   Percent Identity: 41.558

alignment_block:
 US-09-651-563-808 x AAY83169   ..

 Align seg 1/1  to: AAY83169  from: 1  to: 79

     453 AAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGAC 502
         |||      |||      ||||||:::|||   ||||||||||||
       6 LysThrAlaProProIleGluSerGlnAspTyrThrProGlyGln..... 20
```

```
     503 AGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTG 552
         |||:::|||:::   :::   ::::::   |||   ::::::   :::
      21 .GluArgAspGluGlyAlaLeuAspPheGlnValProSerLeuAlaAlaT 37

     553 ATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGGG 602
           |||   ||||||   :::   :::||||||   :::   |||   |||
      37 yrLeuTrpGluLeuThrArgProLysThrGlyGlyGluArgGlyAspGly 53

     603 TTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATGC 652
           :::   :::||||||::: :::|||:::   |||      |||:::|
      54 ProAsnValLysGlyGluSer.LeuProAsnLeuGluProValLysIleP 70

     653 CAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         ||||||||||||||   ||||||:::|||
      70 roGluAlaGlyGluGlyGlnProSerVal 79

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07749

seq_documentation_block:
ID   AAB07749 standard; Protein; 146 AA.
XX
AC   AAB07749;
XX
DT   07-NOV-2000  (first entry)
XX
DE   A human cancer-associated protein-3 (CAP-3).
XX
KW   Human; cancer-associated protein; CAP; CAP-1; CAP-2; CAP-3;
KW   cell proliferation; arteriosclerosis; cirrhosis; psoriasis; hepatitis;
KW   cancer; leukaemia; melanoma; autoimmune disorder; inflammatory disorder;
KW   acquired immunodeficiency syndrome; AIDS; anaemia; asthma;
KW   Crohn's disease; multiple sclerosis; microbial infection.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   19
FT                   /note= "potential casein kinase II phosphorylation site"
FT   Modified-site   23
FT                   /note= "potential casein kinase II phosphorylation site"
FT   Modified-site   41
FT                   /note= "potential casein kinase II phosphorylation site"
FT   Modified-site   71
FT                   /note= "potential casein kinase II phosphorylation site"
FT   Modified-site   81
FT                   /note= "potential protein kinase C phosphorylation site"
FT   Modified-site   113
FT                   /note= "potential casein kinase II phosphorylation site"
FT   Modified-site   136
FT                   /note= "potential casein kinase II phosphorylation site"
XX
PN   WO200043508-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-US01565.
XX
PR   22-JAN-1999;   99US-0183027.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL,  Yue H,  Tang YT,  Azimzai Y;
XX
DR   WPI; 2000-482911/42.
DR   N-PSDB; AAA59322.
XX
PT   Isolated nucleic acids encoding human cancer-associated proteins,
PT   useful for treating cancers and autoimmune/inflammatory disorders -
XX
PS   Claim 1; Fig 3A-B; 89pp; English.
XX
CC   The present sequence represents a human cancer-associated protein (CAP).
CC   The specification describes CAP-1, CAP-2 and CAP-3. CAP polynucleotides
CC   and polypeptides may be used in the prevention, treatment and diagnosis
CC   of diseases associated with inappropriate CAP expression. For example,
CC   they may be used to treat disorders associated with decreased CAP
CC   expression such as disorders of cell proliferation
CC   (e.g. arteriosclerosis, cirrhosis, psoriasis and hepatitis), cancers
CC   (e.g. leukaemia, melanoma and cancers of the breast lung and prostate),
CC   autoimmune/inflammatory disorders (e.g. acquired immunodeficiency
CC   syndrome (AIDS), anaemia, asthma, Crohn's disease and multiple
CC   sclerosis), and microbial infections. The CAP polypeptides may be used
CC   as antigens in the production of antibodies against CAP and in assays
CC   to identify modulators (agonists and antagonists) of CAP expression and
CC   activity.
XX
SQ   Sequence   146 AA;


alignment_scores:
            Quality:  109.00            Length:     96
              Ratio:  1.879               Gaps:      3
 Percent Similarity:  60.417   Percent Identity: 34.375

alignment_block:
 US-09-651-563-808 x AAB07749   ..

 Align seq 1/1  to: AAB07749  from: 1  to: 146

     423 TCCCAGGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAGCTGA 472
         |||:::|||   |||:::   :::|||::::::   |||      :::::
      20 SerGluGluSerSerAspGluGlnProAspGluValGluSerProThrGl 36

     473 AAGTCGGGATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGATACAG 522
         :|||:::|||   ||||||:::::::::   :::|||||||:::   ::::
      36 nSerGlnAspSerThrProAlaGluGlu...ArgGluAspGluGlyAlaS 52

     523 CTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGCAAGAGCTGCATCAG 572
         ::            ::::::   |||:::|||   |||||||||   |||
      52 erAlaAlaGlnGlyGlnGluProGluAlaAspSerGlnGluLeuValGln 68

     573 TCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGA 622
            :::||||||   :::   |||    |||      :::|||
      69 ProLysThrGlyCysGluLeuGlyAsp.GlyProAspThrLysArgValC 85

     623 TAATACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAG... 669
           :::   :::||||||:::   |||:::|||      |||   |||
      85 ysLeuArgAsnGluGluGlnMetLysLeuProAlaGluGlyProGluPro 101

     670 ............CAACCACAAGTTTAAATGAAGACA 693
                     |||   ||||||      ||||||
     102 GluAlaAspSerGlnGluGlnValHisProLysThr 113

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB33006

seq_documentation_block:
ID   AAB33006 standard; Protein; 396 AA.
XX
AC   AAB33006;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Pinus radiata transcription factor protein sequence #133.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB.
XX
OS   Pinus radiata.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
```

```
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 8; Page 389; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is one such
CC   transcription factor. The transcription factor may be used to produce a
CC   plant having modified gene expression such as a woody plant e.g. a
CC   eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or
CC   to modify the activity of a polypeptide in a plant. The transcription
CC   factors of the present invention are members from the following families
CC   of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic
CC   helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain
CC   zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2
CC   Cys2His2, CCAAT box elements and MYB.
XX
SQ   Sequence   396 AA;


alignment_scores:
            Quality:  103.50              Length:    226
              Ratio:  0.855                Gaps:     13
 Percent Similarity:  53.540   Percent Identity: 29.646

alignment_block:
 US-09-651-563-808 x AAB33006   ..

 Align seg 1/1  to: AAB33006  from: 1  to: 396

      63 GACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGA 112
         ||||||:::||||||:::   ||||||||||||   |||         ::
     132 AspThrGlyGlyArgGlyTrpGlyArgSerAsnGlyGluGlu......As 146

     113 GTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGTG 162
         :         ||||||   ||||||                     |||:
     146 pThrGlyGlyArgGlyTrpSerArg.....................SerA 156

     163 ACTCAGAGTGAGAA............GGCCCTCGAAGTCGTCGTCCCTCT 200
         :::::|||   :::            |||   :::||||||      :::
     156 snGlyGluAspAspAlaAlaAlaGlyGlyGlyGlnSerArgGlyArgGly 172

     201 CATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTA 250
         ::::::      |||                          |||::::::
     173 ArgGlyArgGlyArg.......................GlyArgGlyPh 181

     251 GGGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGC 300
          |||||||||   :::   ::::::   |||:::|||::::::
     181 eGlyGlyArgGlySerGlyArgPheGlyGlyGlyGlyAspSer..... 195

     301 TGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATT 350
              |||   :::   :::   :::         |||||||||::::::
     196 ..TyrGlyTyrAspAlaAsnGlyGlnAspArgProProArgGlnGlnPhe 211

     351 CTGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACAC 400
           |||:::      :::|||      ||||||          :::
     212 GluAspThr......AsnThrPheThrGlyThr.........AspAsnTr 223

     401 AAACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCC 450
          :::|||         |||      :::|||   :::|||   ||||||:
     223 pAspThrProGluValSerValValAspGluAlaLysAsnValGluProG 240

     451 AAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAG 500
         ::::::::::|||      |||:::         |||||||||
     240 luGlnLysLysProGluGluGluAla.........ThrProGly...Val 252

     501 ACAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGG 550
         |||:::|||:::::::|||:::   |||         ::::::|||
     253 ThrSerGluAsnLysAspAsnLysGluGluGluAspAsnGluMetThrLe 269

     551 TGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTG 600
          |||      |||    :::   :::            :::|||:::
     269 uAspGluTyrGluLys.LeuLeuAsnGluLysArgLysThrLeuGluAla 285

     601 GGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAAT 650
            ::::::::::::   |||:::||| |||   |||:::      |||
     286 LeuLysAsnAlaGluArgLysValIle.LeuAspArgAspPheGluLysM 302

     651 GCCAGAAGCAGGTGAAGAGCAACCAC.............AAGTTTAAATGA 688
           |||         |||:::|||               ||| ||||||::
     302 etGlnLeuValAspLysLysAsnAspGlyIlePheIleLys.LeuAsnSe 318

     689 AGACAAGCTGAAACAACGCAAG 710
         ::::|||   :::||||||||||
     318 rGluLysGluArgGlnArgLys 325

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW79958

seq_documentation_block:
ID   AAW79958 standard; Protein; 578 AA.
XX
AC   AAW79958;
XX
DT   02-FEB-1999  (first entry)
XX
DE   Human progression elevated gene-2 (PEG-3) protein.
XX
KW   Progression elevated gene-3; PEG-3; human; tumour progression;
KW   DNA repiar; tumorigenesis; angiogenesis; metastasis; melanoma;
KW   brain cancer; cervix cancer; prostate cancer; lung cancer;
KW   colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy.
XX
OS   Homo sapiens.
XX
PN   WO9842315-A1.
XX
PD   01-OCT-1998.
XX
PF   20-MAR-1998;   98WO-US05793.
XX
PR   21-MAR-1997;   97US-0821818.
XX
PA   (GENQ-) GENQUEST INC.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Fisher PB,  Goldstein NI,  Su Z,  Zhang N;
XX
DR   WPI; 1998-557025/47.
DR   N-PSDB; AAV65765.
XX
PT   New isolated Progression Elevated Gene-3  - used to develop products
PT   for e.g. modulating DNA damage and repair pathways, cancer
PT   progression or oncogene mediated transformation and angiogenesis.
XX
PS   Claim 17; Fig 13A-C; 225pp; English.
XX
CC   This polypeptide comprises human progression elevated gene-3 (PEG-3)
CC   protein.  The amino acid sequence was deduced from an isolated cDNA
CC   clone (see AAV65765).  PEG-3 expression (1.5 and 2.8 kb transcripts)
CC   is elevated in most human tumour cell lines.  The amount of PEG-3
CC   in a cell is also an indicator of DNA damage in that cell.  Cells
```

```
CC   in which a reporter gene in under control of the promoter or
CC   regulatory region (see AAV65766) of the rat PEG-3 gene can be used in
CC   methods for identifying agents that modulate PEF-3 expression or the
CC   ability of PEG-3 to induce progression, or for determining whether
CC   an agent is capable of inhibiting DNA damage and repair pathways,
CC   cancer progression or oncogene-mediated transformation. Compounds
CC   that induce DNA damage or which regulate angiogenesis can also be
CC   identified using such cells. Transgenic animals and vaccines
CC   comprising PEG-3 polypeptides and an immune response enhancer are
CC   also claimed. Antibodies (especially monoclonal) to the PEG-3 may
CC   be used to determine tumour progression of a cell. Cells can be
CC   protected from chemotherapeutic damage by inhibiting or eliminating
CC   the expression of PEG-3 in the cells. The methods can be applied
CC   to a progression phenotype comprising anchorage-independent growth,
CC   tumorigenesis, angiogenesis or metastasis, to melanoma, brain,
CC   cervical, prostate, lung or colorectal cancer, neuroblastoma or
CC   glioblastoma (all claimed).
XX
SQ   Sequence   578 AA;


alignment_scores:
           Quality:  103.00            Length:    179
             Ratio:  1.198               Gaps:     11
 Percent Similarity:  48.045   Percent Identity: 27.374

alignment_block:
 US-09-651-563-808 x AAW79958   ..

 Align seg 1/1  to: AAW79958  from: 1  to: 578

     112 AGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGT 161
         ::::::::::::   ||||||:::|||   :::         |||:::
     220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluTy 236

     162 GACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTCATGCGGTGCC 211
                 |||:::      :::                   :::   |
     236 rHisThrArgGluArgProLysGlnGluGlyGluThrLysProGluGlnH 253

     212 ACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAGGGAGGAAGGA 261
         ||      |||   |||                          |||||||||
     253 isArgAlaGlyGlnSerHisProCysGlnAsnAlaGluSerGluGluGly 269

     262 GGGCCGAGG...........AGTGGAGGGGCTCAGGCGAAGCTGGGGTG 299
         ||||||:::           ||||||:::|||      |||
     270 GlyProGluThrSerValCysSerGlySerAlaPheLeuLys........ 283

     300 CTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGAT 349
          :::|||   ||| |||            ||||||      |||||||:::
     284 .AlaTrpValTyr.Arg...........ProGlyGluAspThrGluGlu 295

     350 TCTGGACTCCCCA......GACGGGACCAGGAGAGGGACGGCATGAGCGA 393
         ::: ::::::|||      |||:::::::::::::   ||||||
     296 Glu.GluAspProAspLeuAspSerAlaGluGluAspThrAla....... 309

     394 CACACACAAACACAGAACCACACAGC..................... 419
           ||||||   |||   |||||||:::
     310 ..HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpVal 325

     420 ..............CAGTCCCAGGAGCCCAGTAATGGAGAG...CCCCA 451
                      ::::::::::|||   ::::::|||:::      ::
     326 TyrArgProGlyGluAspThrGluGluGluAspAspGlyAspTrpAspSe 342

     452 AAAAGAAGAACCAGCAGCTGAAAGT...CGGGATCCT............. 485
         :   ||||||   ||||||:::|||         |||
     342 rAlaGluGluAspAlaAlaGlnSerCysThrThrProHisThrSerAlaP 359

     486 ...................ACACCTGGGCAGCAGACAGAAGAAGATCAG 515
                             ||||||:::::::||||||||::::::
     359 heLeuLysAlaTrpValTyrArgProGlyGluAspThrGluGluGluAsp 375

     516 GATACAGCTGAGATCCCAGTGCGCGACATGGAA 548
         |||:::   ::::::         |||   |||

     376 AspSerGluAsnValAlaProValAspSerGlu 386

seq_name: /SIDS1/gcgdata/geneseq/genesegp/AA1999.DAT:AAY41104

seq_documentation_block:
ID   AAY41104 standard; Protein; 578 AA.
XX
AC   AAY41104;
XX
DT   17-JAN-2000  (first entry)
XX
DE   Human progression elevated gene-3 (PEG-3) amino acid sequence.
XX
KW   Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;
KW   cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; human.
XX
OS   Homo sapiens.
XX
PN   WO9949898-A1.
XX
PD   07-OCT-1999.
XX
PF   31-MAR-1999;   99WO-US07199.
XX
PR   31-MAR-1998;   98US-0052753.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Fisher PB;
XX
DR   WPI; 1999-591184/50.
DR   N-PSDB; AAZ23029.
XX
PT   Novel vectors useful for studying the progression of cancer -
XX
PS   Disclosure; Fig 13A-C; 251pp; English.
XX
CC   The invention relates to an inducible progression-elevated gene-3 (PEG-3
CC   gene) regulatory region functionally linked to a gene encoding a product
CC   that causes or may be induced to cause the death or inhibition of cancer
CC   cell growth. A vector of the invention which contains a gene encoding
CC   thymidine kinase or a product which causes the cell to express a
CC   specific antigen can be administered along with gancyclovir or acyclovir,
CC   or an antibody or fragment to the antigen, respectively, to treat cancer
CC   in a subject. The PEG-3 gene is useful for generating new cloning and
CC   expression vectors, transfected cells, and for developing methods for
CC   cultured growth of such cells. The PEG-3 polynucleotide is also useful as
CC   a source of primers and probes to study the progression of cancer, and to
CC   detect the presence of the gene. The present sequence represents the
CC   amino acid sequence of the human PEG-3 protein.
XX
SQ   Sequence   578 AA;


alignment_scores:
           Quality:  103.00            Length:    179
             Ratio:  1.198               Gaps:     11
 Percent Similarity:  48.045   Percent Identity: 27.374

alignment_block:
 US-09-651-563-808 x AAY41104   ..

 Align seg 1/1  to: AAY41104  from: 1  to: 578

     112 AGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGT 161
         ::::::::::::   ||||||:::|||   :::         |||:::
     220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluTy 236

     162 GACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTCATGCGGTGCC 211
                 |||:::      :::                   :::   |
     236 rHisThrArgGluArgProLysGlnGluGlyGluThrLysProGluGlnH 253

     212 ACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAGGGAGGAAGGA 261
```

```
     ||      |||   |||                              |||||||||
 253 isArgAlaGlyGlnSerHisProCysGlnAsnAlaGluSerGluGluGly 269

 262 GGGCCGAGG............AGTGGAGGGGCTCAGGCGAAGCTGGGGTG 299
     ||||||:::            ||||||:::|||      |||
 270 GlyProGluThrSerValCysSerGlySerAlaPheLeuLys........ 283

 300 CTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGAT 349
      :::|||   ||| |||            ||||||      |||||||:::
 284 .AlaTrpValTyr.Arg............ProGlyGluAspThrGluGlu 295

 350 TCTGGACTCCCCA......GACGGGACCAGGAGAGGGACGGCATGAGCGA 393
     ::: ::::::|||      |||:::::::::::::   ||||||
 296 Glu.GluAspProAspLeuAspSerAlaGluGluAspThrAla....... 309

 394 CACACACAAACACAGAACCACACAGC..................... 419
       ||||||   |||   ||||||:::
 310 ..HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpVal 325

 420 ..............CAGTCCCAGGAGCCCAGTAATGGAGAG...CCCCA 451
                  :::::::::|||   :::::::|||:::      ::
 326 TyrArgProGlyGluAspThrGluGluGluAspAspGlyAspTrpAspSe 342

 452 AAAAGAAGAACCAGCAGCTGAAAGT...CGGGATCCT............. 485
     :   ||||||   ||||||:::|||         |||
 342 rAlaGluGluAspAlaAlaGlnSerCysThrThrProHisThrSerAlaP 359

 486 ...................ACACCTGGGCAGCAGACAGAAGAAGATCAG 515
                         ||||||::::::||||||||::::::
 359 heLeuLysAlaTrpValTyrArgProGlyGluAspThrGluGluGluAsp 375

 516 GATACAGCTGAGATCCCAGTGCGCGACATGGAA 548
     |||:::   ::::::         |||   |||
 376 AspSerGluAsnValAlaProValAspSerGlu 386
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY06514

seq_documentation_block:

```
ID   AAY06514 standard; Protein; 578 AA.
XX
AC   AAY06514;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Human progression elevated protein-3 (PEG-3).
XX
KW   Progression elevated protein-3; PEG-3; human; angiogenesis; cancer;
KW   therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO9937776-A1.
XX
PD   29-JUL-1999.
XX
PF   26-JAN-1999;   99WO-US01623.
XX
PR   29-JAN-1998;   98US-0072941.
PR   26-JAN-1998;   98US-0072469.
XX
PA   (GENQ-) GENQUEST INC.
XX
PI   Fisher PB,  Zhang N;
XX
DR   WPI; 1999-458694/38.
DR   N-PSDB; AAX87383.
XX
PT   Modulation of angiogenesis by altering the expression and/or
PT   activity of a progression-associated protein, especially for cancer
PT   treatment
XX
PS   Claim 1; Page 74-76; 81pp; English.
XX
CC   The present sequence represents human progression elevated protein-3
CC   (PEG-3), as deduced from a cDNA clone (see AAX87383) isolated from a
CC   MCF-7 cDNA library.  PEG-3 represents a novel member of the
CC   gadd34/MyD116 gene family.  Human PEG-3 is generally expressed in
CC   cells that are in progression, including most tumour cell lines.  A
CC   claimed method for modulating angiogenesis in an organism comprises
CC   administering an agent that alters expression and/or activity of a
CC   PEG-3 protein, such as an antisense polynucleotide or antibody.
CC   Also claimed is a method for determining whether an agent
CC   modulates angiogenesis, where the candidate agent is present
CC   within a combinatorial small molecule library.  Vaccines and
CC   pharmaceutical compositions comprising such compounds are also
CC   provided and may be used to prevent angiogenesis, especially
CC   related to cancer cell progression.
XX
SQ   Sequence   578 AA;
```

alignment_scores:

```
           Quality:  103.00              Length:    179
             Ratio:  1.198                 Gaps:     11
 Percent Similarity:  48.045   Percent Identity: 27.374
```

alignment_block:

US-09-651-563-808 x AAY06514 ..

Align seg 1/1 to: AAY06514 from: 1 to: 578

```
 112 AGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGT 161
     ::::::::::::   ||||||:::|||   :::          |||:::
 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluTy 236

 162 GACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTCATGCGGTGCC 211
               |||:::      :::                  :::   |
 236 rHisThrArgGluArgProLysGlnGluGlyGluThrLysProGluGlnH 253

 212 ACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAGGGAGGAAGGA 261
     ||      |||   |||                          |||||||||
 253 isArgAlaGlyGlnSerHisProCysGlnAsnAlaGluSerGluGluGly 269

 262 GGGCCGAGG............AGTGGAGGGGCTCAGGCGAAGCTGGGGTG 299
     ||||||:::            ||||||:::|||      |||
 270 GlyProGluThrSerValCysSerGlySerAlaPheLeuLys........ 283

 300 CTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGAT 349
      :::|||   ||| |||            ||||||      |||||||:::
 284 .AlaTrpValTyr.Arg............ProGlyGluAspThrGluGlu 295

 350 TCTGGACTCCCCA......GACGGGACCAGGAGAGGGACGGCATGAGCGA 393
     ::: ::::::|||      |||:::::::::::::   ||||||
 296 Glu.GluAspProAspLeuAspSerAlaGluGluAspThrAla....... 309

 394 CACACACAAACACAGAACCACACAGC..................... 419
       ||||||   |||   ||||||:::
 310 ..HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpVal 325

 420 ..............CAGTCCCAGGAGCCCAGTAATGGAGAG...CCCCA 451
                  :::::::::|||   :::::::|||:::      ::
 326 TyrArgProGlyGluAspThrGluGluGluAspAspGlyAspTrpAspSe 342

 452 AAAAGAAGAACCAGCAGCTGAAAGT...CGGGATCCT............. 485
     :   ||||||   ||||||:::|||         |||
 342 rAlaGluGluAspAlaAlaGlnSerCysThrThrProHisThrSerAlaP 359

 486 ...................ACACCTGGGCAGCAGACAGAAGAAGATCAG 515
                         ||||||::::::||||||||::::::
 359 heLeuLysAlaTrpValTyrArgProGlyGluAspThrGluGluGluAsp 375

 516 GATACAGCTGAGATCCCAGTGCGCGACATGGAA 548
     |||:::   ::::::         |||   |||
 376 AspSerGluAsnValAlaProValAspSerGlu 386
```

```
OM of: US-09-651-563-808 to: Pending_Patents_AA_Main:*   out_format : pfs

Date: Jun 28, 2001 11:22 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09651563/runat_28062001_145611_7597/app_query.fasta_1.854
 -DB=Pending_Patents_AA_Main -QFMT=fastan -SUFFIX=rapm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09651563_@CGN1_1_149 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-651-563-808
Query length: 781
Database: Pending_Patents_AA_Main:*
Database sequences: 1316263
Database length: 176924844
Search time (sec): 133.930000


score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-560-406-809 +   887.00 1338.80  2.0e-66   160
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-589-184-809 +   887.00 1338.80  2.0e-66   160
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-651-563-809 +   887.00 1338.80  2.0e-66   160
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-658-824-809 +   887.00 1338.80  2.0e-66   160
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-671-325-809 +   887.00 1338.80  2.0e-66   160
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-809 +   887.00 1338.80  2.0e-66   160
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-736-457-809 +   887.00 1338.80  2.0e-66   160
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-533-077-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-546-259-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-560-406-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-589-184-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-651-563-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-658-824-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-671-325-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-736-457-791 +   550.00  830.93  5.0e-38   126
/cgn2_6/ptodata/2/paa/US60_COMB.pep:US-60-147-499-5605 +   536.00  805.81  8.2e-37   192
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-1831 +  507.00  768.38  2.0e-34    96
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-736-457-1831 +  507.00  768.38  2.0e-34    96
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-519-642-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-533-077-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-533-077-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-546-259-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-546-259-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-560-406-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-560-406-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-589-184-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-589-184-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-651-563-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-651-563-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-658-824-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-658-824-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-671-325-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-671-325-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-736-457-786 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-736-457-793 +   423.50  640.89  2.2e-27   108
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-1830 +  380.00  579.93  9.4e-24    64
/cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-736-457-1830 +  380.00  579.93  9.4e-24    64
/cgn2_6/ptodata/2/paa/US60_COMB.pep:US-60-197-873-19905 +  251.00  381.91  7.5e-13
/cgn2_6/ptodata/2/paa/US094_COMB.pep:US-09-488-725A-2733 +  251.00  379.53  7.9e-13
/cgn2_6/ptodata/2/paa/US60_COMB.pep:US-60-197-873-25416 +  251.00  379.53  7.9e-13
/cgn2_6/ptodata/2/paa/US60_COMB.pep:US-60-238-331-127 +  251.00  379.53  7.9e-13   11
/cgn2_6/ptodata/2/paa/US60_COMB.pep:US-60-277-380-36 +  251.00  379.53  7.9e-13   111

seq_name: /cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-560-406-809

seq_documentation_block:
; Sequence 809, Application US/09560406
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT: Bangur, Chaitanya S.
;  APPLICANT: Lodes, Michael A.
;  APPLICANT: Fanger, Gary
;  APPLICANT: Vedvick, Tom
;  APPLICANT: Carter, Darrick
;  APPLICANT: Retter, Marc
;  APPLICANT: Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C7
;  CURRENT APPLICATION NUMBER: US/09/560,406
;  CURRENT FILING DATE: 2000-04-27
;  NUMBER OF SEQ ID NOS: 824
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 809
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-560-406-809


alignment_scores:
           Quality:  887.00              Length:    160
             Ratio:  5.544                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-651-563-808 x US-09-560-406-809    ..

 Align seg 1/1  to: US-09-560-406-809  from: 1  to: 160

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 gGluGluGlyGlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysC 34

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysTrpGlyTyrProSerProArgSerThrTrpAsnProAspArgArgPhe 50

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TrpThrProGlnThrGlyProGlyGluGlyArgHisGluArgHisThrGl 67

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 84

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 117

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     117 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 150

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         ||||||||||||||||||||||||||||||
     151 ProGluAlaGlyGluGluGlnProGlnVal 160

seq_name: /cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-589-184-809

seq_documentation_block:
; Sequence 809, Application US/09589184
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C8
;  CURRENT APPLICATION NUMBER: US/09/589,184
;  CURRENT FILING DATE:  2000-06-05
;  NUMBER OF SEQ ID NOS:  827
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 809
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-589-184-809


alignment_scores:
            Quality:  887.00              Length:    160
              Ratio:  5.544                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-651-563-808 x US-09-589-184-809   ..

 Align seg 1/1  to: US-09-589-184-809  from: 1  to: 160

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 gGluGluGlyGlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysC 34

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysTrpGlyTyrProSerProArgSerThrTrpAsnProAspArgArgPhe 50

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TrpThrProGlnThrGlyProGlyGluGlyArgHisGluArgHisThrGl 67

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 84

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 117

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 150

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         ||||||||||||||||||||||||||||||
     151 ProGluAlaGlyGluGluGlnProGlnVal 160

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-651-563-809

seq_documentation_block:
; Sequence 809, Application US/09651563
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C10
;  CURRENT APPLICATION NUMBER: US/09/651,563
;  CURRENT FILING DATE:  2000-08-29
;  NUMBER OF SEQ ID NOS:  1679
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 809
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-651-563-809


alignment_scores:
            Quality:  887.00              Length:    160
              Ratio:  5.544                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-651-563-808 x US-09-651-563-809   ..

 Align seg 1/1  to: US-09-651-563-809  from: 1  to: 160

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 gGluGluGlyGlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysC 34

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysTrpGlyTyrProSerProArgSerThrTrpAsnProAspArgArgPhe 50

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TrpThrProGlnThrGlyProGlyGluGlyArgHisGluArgHisThrGl 67

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 84

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 84 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100

502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 117

552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
117 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134

602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
134 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 150

652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
    |||||||||||||||||||||||||||||||
151 ProGluAlaGlyGluGluGlnProGlnVal 160

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-658-824-809

seq_documentation_block:
; Sequence 809, Application US/09658824
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C11
;  CURRENT APPLICATION NUMBER: US/09/658,824
;  CURRENT FILING DATE:  2000-09-08
;  NUMBER OF SEQ ID NOS:  1788
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 809
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-658-824-809


alignment_scores:
          Quality:  887.00              Length:    160
            Ratio:  5.544                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-651-563-808 x US-09-658-824-809   ..

 Align seg 1/1  to: US-09-658-824-809  from: 1  to: 160

202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17

252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 17 gGluGluGlyGlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysC 34

302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 34 ysTrpGlyTyrProSerProArgSerThrTrpAsnProAspArgArgPhe 50

352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 51 TrpThrProGlnThrGlyProGlyGluGlyArgHisGluArgHisThrGl 67

402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 84

452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 84 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100

502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 117

552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
117 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134

602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
134 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 150

652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
    |||||||||||||||||||||||||||||||
151 ProGluAlaGlyGluGluGlnProGlnVal 160

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-671-325-809

seq_documentation_block:
; Sequence 809, Application US/09671325
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C12
;  CURRENT APPLICATION NUMBER: US/09/671,325
;  CURRENT FILING DATE:  2000-09-26
;  NUMBER OF SEQ ID NOS:  1825
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 809
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-671-325-809


alignment_scores:
          Quality:  887.00              Length:    160
            Ratio:  5.544                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-651-563-808 x US-09-671-325-809   ..

 Align seg 1/1  to: US-09-671-325-809  from: 1  to: 160

202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17

252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 17 gGluGluGlyGlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysC 34

302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 34 ysTrpGlyTyrProSerProArgSerThrTrpAsnProAspArgArgPhe 50
```

```
     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TrpThrProGlnThrGlyProGlyGluGlyArgHisGluArgHisThrGl 67

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 84

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 117

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 150

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||||
     151 ProGluAlaGlyGluGluGlnProGlnVal 160

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-809

seq_documentation_block:
; Sequence 809, Application US/09702705
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C14
;  CURRENT APPLICATION NUMBER: US/09/702,705
;  CURRENT FILING DATE:  2000-10-30
;  NUMBER OF SEQ ID NOS:  1833
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 809
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-702-705-809


alignment_scores:
            Quality:  887.00              Length:    160
              Ratio:  5.544                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-651-563-808 x US-09-702-705-809   ..

 Align seg 1/1  to: US-09-702-705-809  from: 1  to: 160

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
      17 gGluGluGlyGlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysC 34

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysTrpGlyTyrProSerProArgSerThrTrpAsnProAspArgArgPhe 50

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TrpThrProGlnThrGlyProGlyGluGlyArgHisGluArgHisThrGl 67

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 84

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 117

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 150

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||||
     151 ProGluAlaGlyGluGluGlnProGlnVal 160

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-736-457-809

seq_documentation_block:
; Sequence 809, Application US/09736457
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 809
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-736-457-809


alignment_scores:
            Quality:  887.00              Length:    160
              Ratio:  5.544                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-651-563-808 x US-09-736-457-809   ..

 Align seg 1/1  to: US-09-736-457-809  from: 1  to: 160
```

```
    202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17

    252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     17 gGluGluGlyGlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysC 34

    302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     34 ysTrpGlyTyrProSerProArgSerThrTrpAsnProAspArgArgPhe 50

    352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 TrpThrProGlnThrGlyProGlyGluGlyArgHisGluArgHisThrGl 67

    402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 84

    452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     84 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100

    502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 117

    552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    117 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134

    602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    134 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 150

    652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
        |||||||||||||||||||||||||||||
    151 ProGluAlaGlyGluGluGlnProGlnVal 160
```

seq_name: /cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-533-077-791

```
seq_documentation_block:
; Sequence 791, Application US/09533077
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT: Bangur, Chaitanya S.
;  APPLICANT: Algate, Paul A.
;  APPLICANT: Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C5
;  CURRENT APPLICATION NUMBER: US/09/533,077
;  CURRENT FILING DATE: 2000-03-22
;  NUMBER OF SEQ ID NOS: 800
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 791
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-533-077-791


alignment_scores:
            Quality:  550.00              Length:    227
              Ratio:  4.583                Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x US-09-533-077-791   ..

 Align seg 1/1  to: US-09-533-077-791  from: 1  to: 126
```

```
      2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

     52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
        |||||||||||||||||||
     23 rProLeuLeuArgHisGly..............................  29

    102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

     29 .....................................................  29

    152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

     29 .....................................................  29

    202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

     29 .....................................................  29

    252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

     29 .....................................................  29

    302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

     29 .....................................................  29

    352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                           :::  ||||||||
     30 ..................................Gly..HisThrGl 33

    402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50

    452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66

    502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83

    552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

    602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

    652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
        |||||||||||||||||||||||||||||
    117 ProGluAlaGlyGluGluGlnProGlnVal 126
```

seq_name: /cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-546-259-791

```
seq_documentation_block:
; Sequence 791, Application US/09546259
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C6
;  CURRENT APPLICATION NUMBER: US/09/546,259
;  CURRENT FILING DATE:  2000-04-10
;  NUMBER OF SEQ ID NOS:  803
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 791
```

```
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-546-259-791


alignment_scores:
            Quality:  550.00              Length:    227
              Ratio:  4.583                 Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x US-09-546-259-791   ..

 Align seg 1/1  to: US-09-546-259-791  from: 1  to: 126

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         ||||||||||||||||||
      23 rProLeuLeuArgHisGly.............................. 29

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

      29 ..................................................... 29

     152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

      29 ..................................................... 29

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

      29 ..................................................... 29

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      29 ..................................................... 29

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      29 ..................................................... 29

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                              :::  ||||||||
      30 ..................................Gly..HisThrGl 33

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||
     117 ProGluAlaGlyGluGluGlnProGlnVal 126

seq_name: /cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-560-406-791


seq_documentation_block:
; Sequence 791, Application US/09560406
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT: Bangur, Chaitanya S.
;  APPLICANT: Lodes, Michael A.
;  APPLICANT: Fanger, Gary
;  APPLICANT: Vedvick, Tom
;  APPLICANT: Carter, Darrick
;  APPLICANT: Retter, Marc
;  APPLICANT: Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C7
;  CURRENT APPLICATION NUMBER: US/09/560,406
;  CURRENT FILING DATE: 2000-04-27
;  NUMBER OF SEQ ID NOS: 824
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 791
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-560-406-791


alignment_scores:
            Quality:  550.00              Length:    227
              Ratio:  4.583                 Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x US-09-560-406-791   ..

 Align seg 1/1  to: US-09-560-406-791  from: 1  to: 126

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         ||||||||||||||||||
      23 rProLeuLeuArgHisGly.............................. 29

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

      29 ..................................................... 29

     152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

      29 ..................................................... 29

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

      29 ..................................................... 29

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      29 ..................................................... 29

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      29 ..................................................... 29

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                              :::  ||||||||
      30 ..................................Gly..HisThrGl 33

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
      50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||||
     117 ProGluAlaGlyGluGluGlnProGlnVal 126

seq_name: /cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-589-184-791

seq_documentation_block:
; Sequence 791, Application US/09589184
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C8
;  CURRENT APPLICATION NUMBER: US/09/589,184
;  CURRENT FILING DATE:  2000-06-05
;  NUMBER OF SEQ ID NOS:  827
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 791
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-589-184-791


alignment_scores:
            Quality:  550.00            Length:    227
              Ratio:  4.583               Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x US-09-589-184-791   ..

 Align seg 1/1  to: US-09-589-184-791  from: 1  to: 126

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         |||||||||||||||||
      23 rProLeuLeuArgHisGly.............................. 29

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

      29 ..................................................... 29

     152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

      29 ..................................................... 29

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

      29 ..................................................... 29

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      29 ..................................................... 29

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      29 ..................................................... 29

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                                  :::  ||||||||
      30 ..................................Gly..HisThrGl 33

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||||
     117 ProGluAlaGlyGluGluGlnProGlnVal 126

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-651-563-791

seq_documentation_block:
; Sequence 791, Application US/09651563
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C10
;  CURRENT APPLICATION NUMBER: US/09/651,563
;  CURRENT FILING DATE:  2000-08-29
;  NUMBER OF SEQ ID NOS:  1679
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 791
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-651-563-791


alignment_scores:
            Quality:  550.00            Length:    227
              Ratio:  4.583               Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
```

```
US-09-651-563-808 x US-09-651-563-791   ..

Align seg 1/1  to: US-09-651-563-791  from: 1  to: 126

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         |||||||||||||||||
      23 rProLeuLeuArgHisGly.............................. 29

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

      29 ..................................................... 29

     152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

      29 ..................................................... 29

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

      29 ..................................................... 29

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      29 ..................................................... 29

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      29 ..................................................... 29

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                          :::  ||||||||
      30 ....................;.............Gly..HisThrGl 33

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||
     117 ProGluAlaGlyGluGluGlnProGlnVal 126

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-658-824-791

seq_documentation_block:
; Sequence 791, Application US/09658824
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C11
;  CURRENT APPLICATION NUMBER: US/09/658,824
;  CURRENT FILING DATE:  2000-09-08
;  NUMBER OF SEQ ID NOS:  1788
;  SOFTWARE:  PastSEQ for Windows Version 3.0
; SEQ ID NO 791
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-658-824-791


alignment_scores:
           Quality:  550.00              Length:    227
             Ratio:  4.583                Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x US-09-658-824-791   ..

 Align seg 1/1  to: US-09-658-824-791  from: 1  to: 126

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         |||||||||||||||||
      23 rProLeuLeuArgHisGly.............................. 29

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

      29 ..................................................... 29

     152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

      29 ..................................................... 29

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

      29 ..................................................... 29

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      29 ..................................................... 29

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      29 ..................................................... 29

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                          :::  ||||||||
      30 ....................;.............Gly..HisThrGl 33

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100
```

```
     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||
     117 ProGluAlaGlyGluGluGlnProGlnVal 126

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-671-325-791

seq_documentation_block:
; Sequence 791, Application US/09671325
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C12
;  CURRENT APPLICATION NUMBER: US/09/671,325
;  CURRENT FILING DATE:  2000-09-26
;  NUMBER OF SEQ ID NOS:  1825
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 791
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-671-325-791


alignment_scores:
            Quality:  550.00            Length:    227
              Ratio:  4.583              Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x US-09-671-325-791   ..

 Align seg 1/1  to: US-09-671-325-791  from: 1  to: 126

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         |||||||||||||||||
      23 rProLeuLeuArgHisGly..............................  29

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

      29 ..................................................  29

     152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

      29 ..................................................  29

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

      29 ..................................................  29

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      29 ..................................................  29

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      29 ..................................................  29

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                                 :::  ||||||||
      30 .....................................Gly..HisThrGl 33

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||
     117 ProGluAlaGlyGluGluGlnProGlnVal 126

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-791

seq_documentation_block:
; Sequence 791, Application US/09702705
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C14
;  CURRENT APPLICATION NUMBER: US/09/702,705
;  CURRENT FILING DATE:  2000-10-30
;  NUMBER OF SEQ ID NOS:  1833
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 791
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-702-705-791


alignment_scores:
            Quality:  550.00            Length:    227
              Ratio:  4.583              Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x US-09-702-705-791   ..

 Align seg 1/1  to: US-09-702-705-791  from: 1  to: 126

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       7 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 23

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         |||||||||||||||||
```

```
 23 rProLeuLeuArgHisGly.............................. 29

102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

 29 ..................................................... 29

152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

 29 ..................................................... 29

202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

 29 ..................................................... 29

252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

 29 ..................................................... 29

302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

 29 ..................................................... 29

352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                          :::  ||||||||
 30 ..................................Gly..HisThrGl 33

402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50

452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66

502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83

552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
100 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 116

652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!
117 ProGluAlaGlyGluGluGlnProGlnVal 126
```

OM of: US-09-651-563-808 to: SwissProt_39:* out_format : pfs

Date: Jun 28, 2001 11:30 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09651563/runat_28062001_145612_7654/app_query.fasta_1.854
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09651563_@CGN1_1_28 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-651-563-808
Query length: 781
Database: SwissProt_39:*
Database sequences: 93435
Database length: 34255486
Search time (sec): 36.120000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
SwissProt_39:GGC1_HUMAN	+	138.50	200.41	0.0016	102	! O60829 homo sapiens (human)
SwissProt_39:SRCA_RABIT	+	114.00	147.61	0.1548	908	! P13666 oryctolagus cuniculus
SwissProt_39:GGB1_HUMAN	+	110.00	157.60	0.2676	146	! O75459 homo sapiens (human)
SwissProt_39:GRPA_RAT	+	106.00	147.57	0.5743	246	! P08568 rattus norvegicus (rat)
SwissProt_39:G19P_HUMAN	+	106.00	141.09	0.6161	527	! P14314 homo sapiens (human)
SwissProt_39:NCR2_MOUSE	+	103.50	124.44	1.11	2472	! Q9wu42 mus musculus nuclear
SwissProt_39:GRPB_RAT	+	103.00	143.35	0.9827	247	! P08462 rattus norvegicus (rat)
SwissProt_39:CYB_TRYBB	-	103.00	140.08	1.02	363	! P00164 trypanosoma brucei bru
SwissProt_39:DMP1_MOUSE	+	99.50	132.42	1.96	503	! O55188 mus musculus (mouse)
SwissProt_39:NR41_HUMAN	-	99.50	130.94	1.99	598	! P22736 homo sapiens (human)
SwissProt_39:CA2B_HUMAN	+	98.00	119.78	2.88	1736	! P13942 homo sapiens (human)
SwissProt_39:CYB_LEITA	-	97.00	131.52	2.98	371	! P14548 leishmania tarentolae
SwissProt_39:HMPH_CHICK	+	96.50	133.31	3.18	277	! Q05502 gallus gallus (chicke
SwissProt_39:YT35_STRFR	-	96.00	130.67	3.55	348	! P20186 streptomyces fradiae
SwissProt_39:GGE4_HUMAN	+	95.50	139.25	3.51	117	! Q13068 homo sapiens (human)
SwissProt_39:GGE5_HUMAN	+	95.50	139.25	3.51	117	! Q13069 homo sapiens (human)
SwissProt_39:GGE6_HUMAN	+	95.50	139.25	3.51	117	! Q13070 homo sapiens (human)
SwissProt_39:GGE3_HUMAN	+	95.50	139.18	3.51	118	! Q13067 homo sapiens (human)
SwissProt_39:MY16_MOUSE	+	95.00	123.86	4.50	657	! P17564 mus musculus (mouse)
SwissProt_39:YFHA_BORPE	+	92.50	126.50	6.59	320	! P33445 bordetella pertussis
SwissProt_39:IE63_HSV2H	+	92.50	122.50	6.88	512	! P28276 herpes simplex virus
SwissProt_39:GGE2_HUMAN	+	92.00	134.45	6.56	116	! Q13066 homo sapiens (human)
SwissProt_39:TISD_HUMAN	-	92.00	122.32	7.48	482	! P47974 homo sapiens (human)
SwissProt_39:NR41_CANFA	-	92.00	120.48	7.63	598	! P51666 canis familiaris (dog
SwissProt_39:SM6C_RAT	-	92.00	116.45	7.97	960	! Q9wtl3 rattus norvegicus (rat)
SwissProt_39:ABL2_HUMAN	+	92.00	114.68	8.12	1182	! P42684 homo sapiens (human)
SwissProt_39:OTX1_HUMAN	-	91.50	124.25	7.95	354	! P32242 homo sapiens (human)
SwissProt_39:DMP1_RAT	+	89.50	118.71	11.71	489	! P98193 rattus norvegicus (rat)
SwissProt_39:SM6C_MOUSE	-	89.50	113.22	12.43	931	! Q9wtm3 mus musculus (mouse)
SwissProt_39:S24B_ARATH	-	89.00	111.35	13.77	1069	! Q9m081 arabidopsis thaliana
SwissProt_39:GGE7_HUMAN	+	88.50	129.49	12.28	117	! O76087 homo sapiens (human)
SwissProt_39:CST2_HUMAN	+	88.50	115.90	14.22	577	! P33240 homo sapiens (human)
SwissProt_39:NEUM_SERCA	+	88.00	122.64	14.35	241	! Q98987 serinus canaria (cana
SwissProt_39:G19P_BOVIN	+	88.00	115.88	15.44	533	! Q28034 bos taurus (bovine)
SwissProt_39:CIKE_DROME	+	88.00	109.16	16.61	1174	! Q02280 drosophila melanogas
SwissProt_39:RCQ4_HUMAN	+	88.00	108.91	16.65	1208	! O94761 homo sapiens (human)
SwissProt_39:NR41_RAT	-	87.50	114.22	17.06	597	! P22829 rattus norvegicus (rat)
SwissProt_39:YHL1_EBV	+	87.50	113.36	17.22	660	! P03181 epstein-barr virus (str
SwissProt_39:POLS_RUBVH	-	87.50	109.30	18.00	1063	! P21480 rubella virus (vacci
SwissProt_39:POLS_RUBVR	-	87.50	109.30	18.00	1063	! P19725 rubella virus (vacci
SwissProt_39:KF1A_HUMAN	+	87.50	105.36	18.78	1690	! Q12756 homo sapiens (hum
SwissProt_39:PRPH_MESAU	+	87.00	123.59	16.74	183	! P06680 mesocricetus aurat
SwissProt_39:FRA2_RAT	+	87.00	118.65	17.65	327	! P51145 rattus norvegicus (r
SwissProt_39:SDC3_CAEEL	+	87.00	102.61	21.00	2150	! P34706 caenorhabditis el
SwissProt_39:VIE1_MCMVS	+	86.50	112.85	20.40	595	! P11210 murine cytomegalov

seq_name: SwissProt_39:GGC1_HUMAN

seq_documentation_block:
ID GGC1_HUMAN STANDARD; PRT; 102 AA.
AC O60829;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE G ANTIGEN FAMILY C 1 PROTEIN (PROSTATE-ASSOCIATED GENE PROTEIN 4)
DE (PAGE-4) (PAGE-1) (JM27) (GAGE-9).
GN GAGEC1 OR PAGE4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98393718; PubMed=9724777;
RA Brinkmann U., Vasmatzis G., Lee B., Yerushalmi N., Essand M.,
RA Pastan I.;
RT "PAGE-1, an X chromosome-linked GAGE-like gene that is expressed in
RT normal and neoplastic prostate, testis, and uterus.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10757-10762(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Strom T.M., Nyakatura G., Hellebrand H., Drescher B., Rosenthal A.,
RA Meindl A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: UNKNOWN.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN NORMAL MALE AND
CC FEMALE REPRODUCTIVE TISSUES, PROSTATE, TESTIS, FALLOPIAN TUBE,
CC UTERUS, AND PLACENTA, AS WELL AS IN PROSTATE CANCER, TESTICULAR
CC CANCER, AND UTERINE CANCER.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AF275258; AAF88037.1; -.
DR EMBL; AJ005894; CAA06751.1; -.
DR EMBL; AF238380; AAF62541.1; -.
DR MIM; 300287; -.
SQ SEQUENCE 102 AA; 11153 MW; CE5D07AFBF73301B CRC64;

alignment_scores:
Quality: 138.50 Length: 79
Ratio: 2.770 Gaps: 2
Percent Similarity: 63.291 Percent Identity: 43.038

alignment_block:
US-09-651-563-808 x GGC1_HUMAN ..

Align seg 1/1 to: GGC1_HUMAN from: 1 to: 102

```
    441 GGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACC 490
        ||||||   |||:::|||||||||   ::::::::::::|||      ||
     28 GlyGluSerGlnGlnGluGluProProThrAspAsnGlnAspIleGluPr 44

    491 TGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCG 540
        ||||||||:::   |||           |||      |||      |||
     44 oGlyGlnGluArgGlu.........GlyThrProProIleGluGluArgL 58
```

```
     541 ACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAA 590
           :::|||||||||   ||||||:::      ::::::|||   :::::
      58 ysValGluGlyAspCysGlnGluMetAspLeuGluLysThrArgSerGlu 74

     591 TCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAAC 640
           |||    ||||||   ||||||   |||      |||:::  .:::|
      75 ArgGlyAsp.GlySerAspValLysGluLysThrProProAsnProLysH 91

     641 ACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCA 675
         ||:::|||      |||||||||:::   ||||||
      91 isAlaLysThrLysGluAlaGlyAspGlyGlnPro 102
```

seq_name: SwissProt_39:SRCA_RABIT

seq_documentation_block:

```
ID   SRCA_RABIT     STANDARD;      PRT;   908 AA.
AC   P13666;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   SARCALUMENIN PRECURSOR.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89345602; PubMed=2762314;
RA   Leberer E., Charuk J.H.M., Green N.M., Maclennan D.H.;
RT   "Molecular cloning and expression of cDNA encoding a lumenal calcium
RT   binding glycoprotein from sarcoplasmic reticulum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:6047-6051(1989).
RN   [2]
RP   SEQUENCE OF 1-19 AND 458-908 FROM N.A.
RX   MEDLINE=89123480; PubMed=2521635;
RA   Leberer E., Charuk J.H.M., Clarke D.M., Green N.M., Zubrycka-Gaarn E.,
RA   McLennan D.H.;
RT   "Molecular cloning and expression of cDNA encoding the 53,000-dalton
RT   glycoprotein of rabbit skeletal muscle sarcoplasmic reticulum.";
RL   J. Biol. Chem. 264:3484-3493(1989).
CC   -!- FUNCTION: PERHAPS INVOLVED IN THE REGULATION OF CALCIUM TRANSPORT.
CC   -!- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN. ASSOCIATED
CC       THROUGH CA(2+) WITH THE MEMBRANE.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A CALCIUM-BINDING
CC       GLYCOPROTEIN/160 KDA (SHOWN HERE) AND A SECOND GLYCOPROTEIN/53
CC       KDA; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25750; AAA31189.1; -.
DR   EMBL; J04480; AAA60730.1; -.
DR   PIR; A33280; A33280.
DR   PIR; A33312; A33312.
KW   Calcium-binding; Glycoprotein; Signal; Alternative splicing.
FT   SIGNAL        1    19
FT   CHAIN        20   908       160 KDA SARCALUMENIN.
FT   CHAIN       458   908       53 KDA SARCALUMENIN.
FT   DOMAIN       20   457       ACIDIC DOMAIN, PROBABLY BINDS CALCIUM.
FT   CARBOHYD    102   102       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    716   716       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    824   824       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     21   456       MISSING (IN 53 KDA ISOFORM).
FT   CONFLICT    474   474       Q -> E (IN REF. 2).
SQ   SEQUENCE   908 AA;  97920 MW;  A48CAA221AE1418B CRC64;
```

```
alignment_scores:
            Quality:  114.00              Length:    250
              Ratio:  1.018                 Gaps:     10
 Percent Similarity:  44.800   Percent Identity: 23.600

alignment_block:
 US-09-651-563-808 x SRCA_RABIT   ..

 Align seg 1/1  to: SRCA_RABIT  from: 1  to: 908

      22 GACTCGGGTCCCTGAGGTCTGGATTCTTTCTCCGCTACTGAGACACGGCG 71
         |||:::||||||         |||:::   |||:::   |||:::   ::
     288 AspGlyGlyPro.........AspThrValSerValGlyGluSerLeuGl 301

      72 GGTAGGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGT....... 114
         ::::         |||:::      |||:::   |||   |||
     301 uAlaArgGluGlyAlaGluAspGlnGlyLysProSerProSerGlnGluA 318

     115 .................GAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGG 147
                         |||:::      ||||||:::|||   ::::::
     318 laGlyProAlaAlaAlaGluAlaGlyThrGluProGlyGlyThrLysHis 334

     148 GTTGTGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCC 197
                   :::      ::::::::::::      |||      |||||
     335 SerGluGluGluGlyMetGluGlnAspArgProProGluGlyGlnValPr 351

     198 TCTCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAA 247
         |:::||||||            |||   :::                  :
     351 oValMetArgGlnAspGluGlyGlyGluAla.................S 362

     248 CTAGGGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGG 297
         ::   ||||||   |||   ::::::||||||:::::   :::
     362 erSerGluGluGluGlyAspGluGluGlyGlySerGluGluGlu..... 376

     298 TGCTGTTGGGGGTATCCGAGTCCCAGAA.................GCAC 329
                 |||   ||||||    |||                  :::
     377 ......GluGlyAspProSer.GluGluAspSerGlyGluAspSerGlyA 391

     330 CTGGAACCCCGACAGAAGATTCTGGACTCCCCAGACGGGACCAGGAGAGG 379
           |||:::   :::|||::::::|||       ::::::   ::: ||
     391 spGlyAlaSerSerGluGluAlaGlyAlaAlaSerGluGluAlaSer.Gl 407

     380 GACGGCATGAGCGACACACACAAACACAGAACCACAC............. 416
         |||||||          :::   :::|||:::|||
     407 yThrAlaGlyLeuGlyGluGluGluThrGlnProSerThrGluGlyLeuA 424

     417 .................AGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCC 449
                         ||||||:::|||:::   :::      ||||||
     424 spSerGlyProAlaGlySerGlnAlaGlnAspThrGluAlaGluGluPro 440

     450 CAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCA 499
         ::::::   :::   :::   ||||||        :::|||   :::::
     441 GluGluGlyHisGlnGlyProGluSerProIleThrAlaProGlnGluGl 457

     500 GACAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAG 549
         :||||||:::   ::::::      |||:::|||:::||||||
     457 uThrGluAspValSerGlu......GluValProMetArgAsp....... 469

     550 GTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTT 599

     469 ..................................................... 469

     600 GGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAA 649

     469 ..................................................... 469

     650 TGCCAGAAGCAGGTGAAGAGCAACCACAAGTTTAAATGAAGACAAG 695
         ||||||:::   :::   |||      ||||||||||||||
     470 ....ArgSerHisIleGlnLysThrLeuMetLeuAsnGluAspLys 483
```

seq_name: SwissProt_39:GGB1_HUMAN

seq_documentation_block:

```
ID   GGB1_HUMAN     STANDARD;      PRT;   146 AA.
AC   O75459;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   G ANTIGEN FAMILY B 1 PROTEIN (PROSTATE-ASSOCIATED GENE PROTEIN 1)
DE   (PAGE-1) (GAGE-9) (AL5).
GN   GAGEB1 OR PAGE1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98316329; PubMed=9651357;
RA   Chen M.E., Lin S.-H., Chung L.W.K., Sikes R.A.;
RT   "Isolation and characterization of PAGE-1 and GAGE-7: new genes
RT   expressed in the LNCaP prostate cancer progression model that share
RT   homology with melanoma-associated antigens.";
RL   J. Biol. Chem. 273:17618-17625(1998).
CC   -!- FUNCTION: UNKNOWN.
CC   -!- TISSUE SPECIFICITY: ISOLATED FROM PROSTATE CANCER CELL LINES;
CC       EXPRESSION ASSOCIATED WITH PROGRESSION TO ANDROGEN INSENSITIVE
CC       PHENOTYPE. EXPRESSED IN NORMAL TESTIS AND AT LOWER LEVEL IN NORMAL
CC       PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF058989; AAC25990.1; -.
DR   MIM; 300288; -.
KW   Multigene family.
SQ   SEQUENCE   146 AA;  16134 MW;  91ABA1E3B498DCA2 CRC64;


alignment_scores:
           Quality:  110.00            Length:     96
             Ratio:  1.897               Gaps:      3
 Percent Similarity:  60.417   Percent Identity: 34.375

alignment_block:
 US-09-651-563-808 x GGB1_HUMAN   ..

 Align seg 1/1  to: GGB1_HUMAN  from: 1  to: 146

     423 TCCCAGGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAGCTGA 472
         |||:::|||   |||:::   :::|||:::::   |||      :::::
      20 SerGluGluSerSerAspGluGlnProAspGluValGluSerProThrGl 36

     473 AAGTCGGGATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGATACAG 522
         :|||:::|||   ||||||:::::::::   :::||||||:::   ::::
      36 nSerGlnAspSerThrProAlaGluGlu...ArgGluAspGluGlyAlaS 52

     523 CTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGCAAGAGCTGCATCAG 572
         ::           ::::::   |||:::|||   |||||||||   |||
      52 erAlaAlaGlnGlyGlnGluProGluAlaAspSerGlnGluLeuValGln 68

     573 TCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGA 622
            :::||||||   :::   |||    |||      :::|||
      69 ProLysThrGlyCysGluProGlyAsp.GlyProAspThrLysArgValC 85

     623 TAATACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAG... 669
           :::   :::|||||||:::   |||:::|||      |||   |||
      85 ysLeuArgAsnGluGluGlnMetLysLeuProAlaGluGlyProGluPro 101

     670 ............CAACCACAAGTTTAAATGAAGACA 693
                     |||   ||||||      ||||||

     102 GluAlaAspSerGlnGluGlnValHisProLysThr 113

seq_name: SwissProt_39:GRPA_RAT

seq_documentation_block:
ID   GRPA_RAT       STANDARD;      PRT;   246 AA.
AC   P08568;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SUBMANDIBULAR GLAND SECRETORY GLX-RICH PROTEIN CA PRECURSOR (GRP-CA).
GN   GRP-CA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR;
RX   MEDLINE=91139639; PubMed=1995617;
RA   Cooper L.F., Elia D.M., Tabak L.A.;
RT   "Secretagogue-coupled changes in the expression of glutamine/glutamic
RT   acid-rich proteins (GRPs). Isoproterenol induces changes in GRP
RT   transcript expression and changes in isoforms secreted.";
RL   J. Biol. Chem. 266:3532-3539(1991).
RN   [2]
RP   SEQUENCE OF 17-246 FROM N.A.
RX   MEDLINE=87222334; PubMed=2438276;
RA   Mirels L., Bedi G.S., Dickinson D.P., Gross K.W., Tabak L.A.;
RT   "Molecular characterization of glutamic acid/glutamine-rich secretory
RT   proteins from rat submandibular glands.";
RL   J. Biol. Chem. 262:7289-7297(1987).
CC   -!- FUNCTION: GRP PROTEINS HAVE A MARKED AFFINITY FOR HYDROXYAPATITE.
CC       THEY MAY PLAY A ROLE IN THE FORMATION OF THE PROTECTIVE ACQUIRED
CC       PELLICLE AT THE SALIVA-TOOTH INTERFACE.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SUBMANDIBULAR GLAND ACINAR CELLS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58653; AAA41278.1; -.
DR   EMBL; J02730; AAA41276.1; -.
DR   PIR; A29573; A29573.
DR   PIR; A38647; A38647.
KW   Submandibular gland; Saliva; Repeat; Signal; Multigene family.
FT   SIGNAL        1     18
FT   CHAIN        19    246       SUBMANDIBULAR GLAND SECRETORY GLX-RICH
FT                                 PROTEIN CA.
FT   DOMAIN        67    181       5 X 23 AA TANDEM REPEATS.
FT   REPEAT        67     89       1.
FT   REPEAT        90    112       2.
FT   REPEAT       113    135       3.
FT   REPEAT       136    158       4.
FT   REPEAT       159    181       5.
FT   SIMILAR        1    207       IDENTICAL TO RAT GRP-CB (P08462).
FT   CONFLICT      63     63       A -> P (IN REF. 2).
FT   CONFLICT     135    135       D -> N (IN REF. 2).
FT   CONFLICT     137    137       Q -> E (IN REF. 2).
FT   CONFLICT     208    208       E -> D (IN REF. 2).
FT   CONFLICT     230    230       R -> P (IN REF. 2).
SQ   SEQUENCE   246 AA;  26528 MW;  1C7144CFE67375AA CRC64;


alignment_scores:
           Quality:  106.00            Length:    132
             Ratio:  1.413               Gaps:      5
 Percent Similarity:  56.818   Percent Identity: 31.061
```

```
alignment_block:
 US-09-651-563-808 x GRPA_RAT   ..

 Align seg 1/1  to: GRPA_RAT  from: 1  to: 246

     306 GGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTCTGGA 355
         |||   :::|||:::|||:::::   ||||||   |||          ::
      74 GlySerGluGluGluGlnGlnGlnGlnGluProThrGlnAlaGluAsnGl 90

     356 CTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACAAACA 405
         ::::|||   :::|||                                 :::::::
      90 nGluProProAlaThr.........................SerGlyS 98

     406 CAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCAAAAA 455
         ::|||   ::::::|||:::|||||||||:::::::::|||   |||
      98 erGluGluGluGlnGlnGlnGlnGluProThrGlnAlaGluAsnGln... 113

     456 GAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAGA 505
           ||||||   |||   |||   :::             ||||||   ||
     114 ...GluProProAlaThrSerGlySerGluGluGluGlnGlnGlnGlnGl 129

     506 A............GAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACA 543
         |            |||||||||:::                   :::
     129 uProThrGlnAlaGluAspGlnGlnProProAlaThrSerGlySerGluG 146

     544 TGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACC.........GGG 584
           |||   :::   ||||||      |||:::::::::         :::
     146 luGluGlnGlnGlnGlnGluSerThrGlnAlaGluAsnGlnGluProSer 162

     585 GATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAG 634
         |||:::::::|||   |||   :::   |||   |||:::::::::    |
     163 AspSerAlaGlyGluGlyGlnGluThrGlnProGluGluGlyAsn.ValG 179

     635 AGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCACAA 678
         ||:::      :::   ||||||   ::::::||||||||||||
     179 luSerProProSerSerProGluAsnSerGlnGluGlnProGln 193

seq_name: SwissProt_39:G19P_HUMAN

seq_documentation_block:
ID   G19P_HUMAN     STANDARD;      PRT;   527 AA.
AC   P14314;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROTEIN KINASE C SUBSTRATE, 80 KDA PROTEIN, HEAVY CHAIN (PKCSH)
DE   (80K-H PROTEIN).
GN   PRKCSH OR G19P1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90007553; PubMed=2793184;
RA   Sakai K., Masamichi H., Minoshima S., Kudoh J., Fukuyama R.,
RA   Shimizu N.;
RT   "Isolation of cDNAs encoding a substrate for protein kinase C:
RT   nucleotide sequence and chromosomal mapping of the gene for a human
RT   80K protein.";
RL   Genomics 5:309-315(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97196775; PubMed=9043864;
RA   Ophoff R.A., Terwindt G.M., Vergouwe M.N., van Eijk R.,
RA   Mohrenweiser H., Litt M., Hofker M.H., Haan J., Ferrari M.D.,
RA   Frants R.R.;
RT   "A 3-Mb region for the familial hemiplegic migraine locus on 19p13.1-
RT   p13.2: exclusion of PRKCSH as a candidate gene.";
RL   Eur. J. Hum. Genet. 4:321-328(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03075; AAA52493.1; -.
DR   EMBL; U50326; AAA98668.1; -.
DR   EMBL; U50317; AAA98668.1; JOINED.
DR   EMBL; U50318; AAA98668.1; JOINED.
DR   EMBL; U50319; AAA98668.1; JOINED.
DR   EMBL; U50320; AAA98668.1; JOINED.
DR   EMBL; U50321; AAA98668.1; JOINED.
DR   EMBL; U50322; AAA98668.1; JOINED.
DR   EMBL; U50323; AAA98668.1; JOINED.
DR   EMBL; U50324; AAA98668.1; JOINED.
DR   EMBL; U50325; AAA98668.1; JOINED.
DR   PIR; A32469; A32469.
DR   HSSP; P06705; 1AUI.
DR   MIM; 177060; -.
DR   InterPro; IPR000886; -.
DR   InterPro; IPR002048; -.
DR   PROSITE; PS00014; ER_TARGET; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
KW   Phosphorylation.
FT   DOMAIN      313    336       GLU-RICH (ACIDIC).
FT   MOD_RES      89     89       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES     382    382       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES     389    389       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES     433    433       PHOSPHORYLATION (BY PKC) (POTENTIAL).
SQ   SEQUENCE   527 AA;  59296 MW;  2BB2F9AC14E3FC94 CRC64;


alignment_scores:
           Quality:  106.00              Length:    212
             Ratio:  1.000                 Gaps:     10
 Percent Similarity:  50.000   Percent Identity: 24.528

alignment_block:
 US-09-651-563-808 x G19P_HUMAN   ..

 Align seg 1/1  to: G19P_HUMAN  from: 1  to: 527

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         |||:::||||||      ::::::   |||:::   |||      |||::
     230 SerValThrGluLeu...GlnThrHisProGluLeuAspThrAspGlyAs 245

     102 TTGAAGTGTGAGTGAGAGTGAA..............................  123
         :   ::::::||||||:::|||
     245 pGlyAlaLeuSerGluAlaGluAlaGlnAlaLeuLeuSerGlyAspThrG 262

     124 ..GAGGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGTGACTCAGAGT 171
           :::   ::::::|||         ||||||:::::: :::|||
     262 lnThrAspAlaThrSerPheTyrAspArgValTrpAlaAla.IleArgAs 278

     172 GAGAAGGCCCTCGAAGTCGTCGTCCCTCTCATGCGGTGCCACGCCCATGG 221
          :::      |||:::                        |||   |||   |
     278 pLysTyrArgSerGlu.....................AlaLeuProThrA 288

     222 ACCTTCTTGTCTCGTCACGGCCATAACTAGGGAGG...AAGGAGGGCCGA 268
         |||||   :::   |||   |||   |||   :::   ||||||   :::
     288 spLeuProAlaProSerAlaProAspLeuThrGluProLysGluGluGln 304

     269 GGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGGGGGTATCCGAGT 318

     304 ..................................................  304

     319 CCCAGAAGCACCTGGAACCCCGACAGAAGATTCTGGACTCCCCAGACGGG 368
          |||   :::|||::::::||||||||:::::          :::::::
     305 .ProProValProSerSerProThrGluGluGlu.........GluGluG 318

     369 ACCAGGAGAGGGACGGCATGAGCGACACACACAAACACAGAACCACACAG 418
```

```
      :::::|||::::::      ::::::   ::::::::: |||      ||
  318 luGluGluGluGluGluGluAlaGluGluGluGlu.GluGluAspSe 334

  419 CCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAG 468
      |::::::      |||      :::   ||||||      :::|||||||
  334 rGluGluAlaProProProLeuSerProProGlnProAlaSerProAlaG 351

  469 CTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGAT 518
        |||::::::   |||         :::|||||||:::         |||
  351 luGluAspLysMetProProTyrAspGluGlnThrGlnAlaPheIleAsp 367

  519 ACA............................GCTGAGATCCCAGTGCG 538
      :::                            ||||||      :::::
  368 AlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGluArgSerLeuLy 384

  539 CGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATA 588
      :|||||||||   :::::::::::::|||:::|||          :
  384 sAspMetGluGluSerIleArgAsnLeuGluGln.............G 396

  589 AATCTGGATTTGGGTTCCGGCGTCAAGGTGAA 620
      ::   :::|||   |||      :::||||||
  396 luIleSerPheAspPheGlyProAsnGlyGlu 406

seq_name: SwissProt_39:NCR2_MOUSE

seq_documentation_block:
ID   NCR2_MOUSE     STANDARD;      PRT;  2472 AA.
AC   Q9WU42; Q9WU43; Q9WUC1;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   NUCLEAR RECEPTOR CO-REPRESSOR 2 (N-COR2) (SILENCING MEDIATOR OF
DE   RETINOIC ACID AND THYROID HORMONE RECEPTOR) (SMRT) (SMRTE) (THYROID-,
DE   RETINOIC-ACID-RECEPTOR-ASSOCIATED CO-REPRESSOR) (T3 RECEPTOR-
DE   ASSOCIATING FACTOR) (TRAC).
GN   NCOR2 OR SMRT.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA AND ISOFORM BETA).
RC   TISSUE=Spleen, and Brain;
RX   MEDLINE=99178941; PubMed=10077563;
RA   Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
RT   "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC   TISSUE=Embryo;
RX   MEDLINE=99199215; PubMed=10097068;
RA   Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;
RT   "SMRTe, a silencing mediator for retinoid and thyroid hormone
RT   receptors-extended isoform that is more related to the nuclear
RT   receptor corepressor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
CC   -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC       NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC       PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC   -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC       AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC       WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE
CC       ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIB.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS. ALSO WIDELY EXPRESSED IN EARLY
CC       EMBRYOS.
CC   -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC       ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
CC       AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
CC       INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
CC       DOMAINS (ID1 AND ID2).
CC   -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
CC       SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
CC       SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
CC       FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
CC       SPECIFICITY.
CC   -!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).
CC   -!- SIMILARITY: CONTAINS 2 CORNR BOX.
CC   -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF113001; AAD20944.1; -.
DR   EMBL; AF113002; AAD20945.1; -.
DR   EMBL; AF125671; AAD22972.1; -.
DR   MGD; MGI:1337080; Ncor2.
DR   InterPro; IPR001005; -.
DR   Pfam; PF00249; myb_DNA-binding; 2.
DR   PROSITE; PS50090; MYB_3; 1.
KW   Nuclear protein; Transcription regulation; DNA-binding; Repressor;
KW   Coiled coil; Alternative splicing.
FT   DOMAIN      165    207       COILED COIL (POTENTIAL).
FT   DOMAIN      254    312       INTERACTION WITH SIN3A/B (BY SIMILARITY).
FT   DNA_BIND    429    474       SANT-A (POTENTIAL).
FT   DNA_BIND    608    653       SANT-B (POTENTIAL).
FT   DOMAIN      492    560       COILED COIL (POTENTIAL).
FT   DOMAIN      652    682       COILED COIL (POTENTIAL).
FT   DOMAIN      775    804       PRO-RICH.
FT   DOMAIN      989    999       PRO-RICH.
FT   DOMAIN     1351   1357       PRO-RICH.
FT   DOMAIN     2094   2098       CORNR BOX OF ID1.
FT   DOMAIN     2296   2300       CORNR BOX OF ID2.
FT   DOMAIN      494    507       POLY-GLN.
FT   DOMAIN     1615   1619       POLY-ALA.
FT   DOMAIN     2434   2437       POLY-PRO.
FT   VARSPLIC     36    254       MISSING (IN ISOFORM BETA).
FT   CONFLICT    176    176       M -> RL (IN REF. 2).
FT   CONFLICT    396    402       PPMLYDA -> RHVVRR (IN REF. 2).
FT   CONFLICT    555    555       D -> H (IN AAD20944).
FT   CONFLICT    756    756       T -> M (IN AAD20944).
FT   CONFLICT    785    785       V -> A (IN REF. 2).
FT   CONFLICT    806    846       HHLPHPRLLWTRMNKKPRLLQLPRQRMPRSRSLRPRRSMWE
FT                                -> PSPAAPPATVDKDEQEAPAAPAPQTEDAKEQKSEAEEI
FT                                DVG (IN REF. 2).
FT   CONFLICT    856    856       E -> K (IN AAD20945).
FT   CONFLICT    859    859       E -> K (IN AAD20945).
FT   CONFLICT    867    867       E -> K (IN AAD20945).
FT   CONFLICT    895    895       E -> K (IN AAD20945).
FT   CONFLICT    916    916       S -> F (IN AAD20944).
FT   CONFLICT    975    975       I -> IQ (IN AAD20944).
FT   CONFLICT   1046   1063       PKLPTEPPRWSSGLPFPI -> QSYRLSPHAGHRLPSH
FT                                (IN REF. 2).
FT   CONFLICT   1073   1080       PHAADPSA -> TRADPL (IN REF. 2).
FT   CONFLICT   1133   1133       MISSING (IN REF. 2).
FT   CONFLICT   1149   1149       MISSING (IN REF. 2).
FT   CONFLICT   1157   1157       G -> E (IN REF. 2).
FT   CONFLICT   1172   1201       GSATSGSITKGLPSTRAADGPSYRGSITHG -> APPPVEA
FT                                SPRASQYPGCRRPQLQRLYHPR (IN REF. 2).
FT   CONFLICT   1696   1696       A -> S (IN REF. 2).
FT   CONFLICT   1855   1857       MISSING (IN REF. 2).
FT   CONFLICT   1909   1909       A -> P (IN REF. 2).
FT   CONFLICT   1913   1913       A -> G (IN REF. 2).
FT   CONFLICT   1923   1923       G -> A (IN REF. 2).
FT   CONFLICT   1956   1956       N -> S (IN REF. 2).
FT   CONFLICT   1968   1968       A -> G (IN REF. 2).
FT   CONFLICT   2195   2196       TA -> AV (IN REF. 2).
FT   CONFLICT   2213   2214       LE -> SK (IN REF. 2).
```

```
FT   CONFLICT    2224   2224       T -> A (IN REF. 2).
SQ   SEQUENCE   2472 AA;  270856 MW;  2A58F4DF7B79285B CRC64;


alignment_scores:
            Quality:  103.50              Length:    279
              Ratio:  0.815                Gaps:     13
 Percent Similarity:  45.520   Percent Identity: 24.014

alignment_block:
 US-09-651-563-808 x NCR2_MOUSE   ..

 Align seg 1/1  to: NCR2_MOUSE  from: 1  to: 2472

      25 TCGGGTCCCTGAGGTCTGGATTCTTTCTCCGCTACTGAG..... 63
         |||||||||   ::::::::::::   |||   ||||||
     734 SerGlyProAlaAlaValAsnAsnSerSerAspThrGluSerValProSe 750

      64 ..............ACACGGCGGGTAGGTCCACAGGCAGATCCAACTG 97
                        |||:::   :::|||||||:::      |||||||
     750 rProArgSerGluAlaThrLysAspThrGlyProLys......ProThrG 765

      98 GGAGTTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTT.CCGGAG 146
         ||:::   ::::::      :::   :::   |||:::      ||||||
     765 lyThrGluAlaLeuProAlaAlaThrGlnProProValProProProGlu 781

     147 GGTTGTGTGGTCAGTGACTCAGAG.................TGAGAAGG 178
               :::|||:::   :::|||                  :::||
     782 GluProAlaValAlaProAlaGluProSerProValProAspAlaSerGl 798

     179 CCCTCGAAGTCGTCGTCCCTCTCATGCGGTGCCACGCCCA.........T 219
         ||||   |||   :::|||||||||   :::|||:::|||         |
     798 yProProSerProGluProSerHisHisLeuProHisProArgLeuLeuT 815

     220 GGACCTTCTTGTCTCGTCACGGCCATAACTAGGGAGGAAGGAGGGCCGAG 269
         |||||   :::::::::      :::
     815 rpThrArgMetAsnLysLysProArg........................ 823

     270 GAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGGGGGTATCCGAGTC 319
                                  ||||||   :::   :::::
     824 ..........................LeuLeuGlnLeuProArgGl 830

     320 CCAG............AAGCACCTGGAACCCCGACAGAAGATTCTGGACT 357
         ::::            :::   |||:::||||||:::::::::   ::::
     830 nArgMetProArgSerArgSerLeuArgProArgArgSerMetTrpGluL 847

     358 CCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACAAACACA 407
         ::|||:::
     847 ysProGlu.......................................... 849

     408 GAACCACACAGCCAGTCCCAGGAGCCC...AGTAATGGAGAGCCCCAAAA 454
         ||||||:::::::::::::::   |||   |||      :::   :::||
     850 GluProGluAlaSerGluGluProProGluSerValLysSerAspHisLy 866

     455 AGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAG 504
         |||||||         |||   :::|||   :::   |||      ||||
     866 sGluGluThrGluGluGluProGluAspLysAlaLysGly......ThrG 881

     505 AAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAA...... 548
         ||      :::   ::::::|||   |||::::::      |||
     881 luAlaIleGluThrValSerGluAlaProLeuLysValGluGluAlaGly 897

     549 ..............................GGTGATCTGCAAGAGCT 565
                                       |||      |||:::
     898 SerLysAlaAlaValThrLysGlySerSerSerGlyAlaThrGlnAspSe 914

     566 GCATCAGTCAAACACC..............................G 582
             :::|||   |||  .                          |
     914 rAspSerSerAlaThrCysSerAlaAspGluValAspGluProGluGlyG 931

     583 GGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAA 632
         ||||||||                        |||:::      :::
     931 lyAspLys.....................GlyArgLeuLeuSerPro 939

     633 AGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCACAAGT.. 680
         |||   :::|||   :::      :::      ||||||:::||||||
     940 ArgProSerLeuLeuThrProAlaGlyAspProArgAlaSerThrSerPr 956

     681 .........TTAAATGAAGACAAGCTGAAACAACGC 707
                   |||:::      :::||||||||||||
     956 oGlnLysProLeuAspLeuLysGlnLeuLysGlnArg 968

seq_name: SwissProt_39:GRPB_RAT

seq_documentation_block:
ID   GRPB_RAT       STANDARD;      PRT;   247 AA.
AC   P08462;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SUBMANDIBULAR GLAND SECRETORY GLX-RICH PROTEIN CB PRECURSOR (GRP-CB)
DE   (CONTIGUOUS REPEAT POLYPEPTIDE) (CRP).
GN   GRP-CB.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=87165965; PubMed=3558393;
RA   Heinrich G., Habener J.F.;
RT   "Genes encoding proteins with homologous contiguous repeat sequences
RT   are highly expressed in the serous cells of the rat submandibular
RT   gland.";
RL   J. Biol. Chem. 262:5262-5270(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR;
RX   MEDLINE=91139639; PubMed=1995617;
RA   Cooper L.F., Elia D.M., Tabak L.A.;
RT   "Secretagogue-coupled changes in the expression of glutamine/glutamic
RT   acid-rich proteins (GRPs). Isoproterenol induces changes in GRP
RT   transcript expression and changes in isoforms secreted.";
RL   J. Biol. Chem. 266:3532-3539(1991).
CC   -!- FUNCTION: GRP PROTEINS HAVE A MARKED AFFINITY FOR HYDROXYAPATITE.
CC       THEY MAY PLAY A ROLE IN THE FORMATION OF THE PROTECTIVE ACQUIRED
CC       PELLICLE AT THE SALIVA-TOOTH INTERFACE.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SUBMANDIBULAR GLAND ACINAR CELLS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31032; AAA40969.1; -.
DR   EMBL; M58654; AAA41279.1; ALT_SEQ.
DR   PIR; A29545; A29545.
DR   PIR; B38647; B38647.
KW   Submandibular gland; Saliva; Repeat; Signal; Multigene family.
FT   SIGNAL        1     18
FT   CHAIN        19    247       SUBMANDIBULAR GLAND SECRETORY GLX-RICH
FT                                 PROTEIN CB.
FT   DOMAIN       67    181       5 X 23 AA TANDEM REPEATS.
FT   REPEAT       67     89       1.
FT   REPEAT       90    112       2.
FT   REPEAT      113    135       3.
FT   REPEAT      136    158       4.
FT   REPEAT      159    181       5.
FT   SIMILAR       1    207       IDENTICAL TO RAT GRP-CA (P08568).
SQ   SEQUENCE   247 AA;  26948 MW;  72A5075BCCC314E4 CRC64;
```

```
alignment_scores:
           Quality:  103.00            Length:    139
             Ratio:  1.304               Gaps:      7
 Percent Similarity:  56.835   Percent Identity: 28.058

alignment_block:
 US-09-651-563-808 x GRPB_RAT   ..

 Align seg 1/1  to: GRPB_RAT  from: 1  to: 247

     306 GGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTCTGGA 355
         |||   :::|||:::|||:::::   ||||||   |||         ::
      74 GlySerGluGluGluGlnGlnGlnGlnGluProThrGlnAlaGluAsnGl 90

     356 CTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACACAAACA 405
         ::::|||   :::|||                                :::::::
      90 nGluProProAlaThr.........................SerGlyS 98

     406 CAGAACCACACAGCCAGTCCCAGGAGCCCAGT.........AATGGAGAG 446
         ::|||   :::::::|||:::||||||||||:::         |||   |||
      98 erGluGluGluGlnGlnGlnGlnGluProThrGlnAlaGluAsnGlnGlu 114

     447 CCC.............................CAAAAAGAAGAACCAGC 466
         |||                             |||::::::::::|||::
     115 ProProAlaThrSerGlySerGluGluGluGlnGlnGlnGlnGlnProTh 131

     467 A...GCTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAGAAGAAGATC 513
         :   ||||||:::::::::|||         :::   :::|||||||:::|
     131 rGlnAlaGluAsnGlnGluProProAlaThrSerGlySerGluGluGluG 148

     514 AGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGCAAGAG 563
         ||:::      |||            :::   :::::::::   ||||||
     148 lnGlnGlnGlnGlu............SerThrGlnAlaGluAsnGlnGlu 160

     564 CTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCA 613
                       :::|||:::::::|||   |||   :::   ||
     161 ..............ProSerAspSerAlaGlyGluGlyGlnGluThrGl 172

     614 AGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGT 663
         |   |||:::::::::    |||:::      :::   ||||||   :::
     172 nProGluGluGlyAsn.ValGluSerProProSerSerProGluAsnSer 188

     664 GAAGAGCAACCACAA 678
         :::||||||||||||
     189 GlnGluGlnProGln 193

seq_name: SwissProt_39:CYB_TRYBB

seq_documentation_block:
ID   CYB_TRYBB      STANDARD;      PRT;   363 AA.
AC   P00164;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME B.
GN   COB OR CYTB.
OS   Trypanosoma brucei brucei.
OG   Mitochondrion.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A. (KINETOPLAST).
RX   MEDLINE=84041494; PubMed=6314266;
RA   Benne R., de Vries B.F., van den Burg J., Klaver B.;
RT   "The nucleotide sequence of a segment of Trypanosoma brucei
RT   mitochondrial maxi-circle DNA that contains the gene for
RT   apocytochrome b and some unusual unassigned reading frames.";
RL   Nucleic Acids Res. 11:6925-6941(1983).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X00017; CAA24915.1; -.
DR   PIR; A00160; CBUTB.
DR   InterPro; IPR000179; -.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme; Kinetoplast.
FT   METAL        74     74       IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL        88     88       IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL       175    175       IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL       189    189       IRON 1 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   363 AA;  43455 MW;  2A37CD827E55923B CRC64;


alignment_scores:
           Quality:  103.00            Length:    191
             Ratio:  1.288               Gaps:     12
 Percent Similarity:  41.885   Percent Identity: 23.560

alignment_block:
 US-09-651-563-808/rev x CYB_TRYBB   ..

 Align seg 1/1  to: CYB_TRYBB  from: 1  to: 363

     705 GTTGTTTCAGCTTGTCTTCATTTAAACTTGTGGTTGCTCTTCACCTGCTT 656
         :::   :::::|||||           |||||||:::|||:::|||||
      37 IleCysGlyValCysLeu.........AlaTrpLeuPhePheSerCysPh 50

     655 C............TGGCATTTTACAGTGTTCCTCTTTAGGTATTATCTTC 618
         |            |||:::|||::::::|||||:::   :::
      50 eIleCysSerAsnTrpTyrPheValLeuPheLeuTrpAspPhe....... 64

     617 ACCTTGACGCCGGAACCCAAATCCAGATTTATCCCCGGTGTTTGACTGAT 568
           |||   |||      |||:::           |||:::|||:::
      65 .....AspLeuGlyPheValIleArgSerValHisIleCysPheThrSer 79

     567 GCAGCTCTTGCAGATCACCTTC........................... 546
           |||      :::   :::
      80 Leu.LeuTyrLeuLeuLeuTyrIleHisIlePheLysSerIleThrLeuI 96

     545 .................CATGTCGCGCACTGGGATCTCAGCTGTATCCTG 513
                         |||:::      |||   ::::::   ||||||
      96 leIleLeuPheAspThrHisIleLeuValTrpPheIleGlyPheIleLeu 112

     512 ...ATCTTCTTCTGTCTGCTGCCCAGGTGTAGGATCCCGACTTTCAGCTG 466
            :::|||:::   ::::::      :::|||      |||   ::::
     113 PheValPheIleIleIleIle.AlaPheIleGlyTyrValLeuProCysT 129

     465 CTGGTTCTTCTTTTTGGGGC........................... 446
         ::      |||:::||||||
     129 hrMetMetSerTyrTrpGlyLeuThrValPheSerAsnIleIleAlaThr 145

     445 TCTCCATTACTGGGCTCCTGG......GACTGGCTGTGTGGTTCTGTGTT 402
            |||:::||||||   |||        |||:::   ||||||   ||
     146 ValProIleLeuGlyIleTrpLeuCysTyrTrpIleTrpGlySerGluPh 162

     401 TGTG..............TGTGTCGCTCATGCCGTCCCTCTCCTGGTCC 367
```

```
      |:::                          |||:::::::   :::|||:::|
    162 eIleAsnAspPheThrLeuLeuLysLeuHisValLeuHisValLeuLeuP 179

    366 CG..................................................... 365
        ||
    179 roPheIleLeuLeuIleIleLeuIleLeuHisLeuPheCysLeuHisTyr 195

    364 TCTGGGGAGTCCAGAATCTTCTGT.....................CG 339
            :::|||      ||||||                      ||
    196 PheMetSerSerAspAlaPheCysAspArgPheAlaPheTyrCysGluAr 212

    338 GGGTTCCAGGTGCTTCTGG 320
        |   |||   |||:::|||
    212 gLeuSerPheCysMetTrp 218
```

seq_name: SwissProt_39:DMP1_MOUSE

seq_documentation_block:

```
ID   DMP1_MOUSE      STANDARD;      PRT;   503 AA.
AC   O55188;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1 PRECURSOR (DENTIN MATRIX
DE   PROTEIN-1) (DMP-1) (AG1).
GN   DMP1 OR DMP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER; TISSUE=Molar;
RX   MEDLINE=98184255; PubMed=9525343;
RA   McDougall M., Gu T.T., Luan X., Simmons D., Chen J.;
RT   "Identification of a novel isoform of mouse dentin matrix protein 1:
RT   spatial expression in mineralized tissues.";
RL   J. Bone Miner. Res. 13:422-431(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ; TISSUE=Spleen;
RA   Feng J.Q., Traianedes K., Luan X., McDougall M.;
RT   "Study of murine Dmp-1 gene function and regulation.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COULD BE INVOLVED IN THE MINERALIZATION OF EXTRACELLULAR
CC       MATRIX.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICURLARLY IN
CC       ODONTOBLAST, AMELOBLAST AND CEMENTOBLAST. ALSO EXPRESSED IN BONE
CC       PARTICULARLY IN OSTEOBLAST.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U65020; AAB93764.1; -.
DR   EMBL; AJ242625; CAB59629.1; -.
DR   MGD; MGI:94910; Dmp1.
KW   Extracellular matrix; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    503       DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT   DOMAIN       41     44       POLY-PRO.
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    394    394       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        350    352       CELL ATTACHMENT SITE (POTENTIAL).
FT   CONFLICT     35     35       D -> G (IN REF. 2).
FT   CONFLICT     67     67       H -> D (IN REF. 2).
FT   CONFLICT     99     99       D -> E (IN REF. 2).
FT   CONFLICT    116    116       G -> Y (IN REF. 2).
FT   CONFLICT    137    137       A -> T (IN REF. 2).
SQ   SEQUENCE   503 AA;  54000 MW;  9E8AF9F2729F113A CRC64;
```

alignment_scores:

```
            Quality:   99.50            Length:    281
              Ratio:  0.809              Gaps:     12
 Percent Similarity:  43.772   Percent Identity: 23.132
```

alignment_block:

US-09-651-563-808 x DMP1_MOUSE ..

Align seg 1/1 to: DMP1_MOUSE from: 1 to: 503

```
     22 GACTCGGGTCCCTGAGGTCTGGATTCTTTCTCCGCTACTGAGACACGGCG 71
        |||||||||            |||:::      :::   |||   :::
    104 AspSerGly............AspAspThrPheGlyAspGluAspAsnGl 116

     72 GGTAGGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGTGAGAGTG 121
         :::||||||:::              |||:::   |||      :::|||:
    116 yLeuGlyProGluGluGlyGlnTrpGlyGlyProSerLysLeuAspSerA 133

    122 AAGAGGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGTGACTCAGAGT 171
        ::|||:::   |||                    ::::::||||||||
    133 spGluAspSerAla.....................AspThrThrGlnSer 142

    172 GAGAAGGCCCTCGAAGTCGTCGTCCCTCTCATGCGGTGCCACGCCCATGG 221
        ::::::    |||   |||::::::   |||:::      ||||||   |
    143 SerGlu.AspSerThrSerGlnGluAsnSerAlaGlnAspThrProSerA 159

    222 ACCTTCTT.....................GTCTCGTCACGGCCATAACTA 250
        ||                          :::   |||||||||
    159 spSerLysAspGlnAspSerGluAspAspAlaHisSerArgProAspAla 175

    251 GGG.............................AGGAAGGAGGGCCGAGG 270
        |||                           ::::::   |||   ||
    176 GlyAspSerAlaGlnHisSerGluSerGluGluGlnArgValGlyGlyGl 192

    271 AGTGGAGGGGCTCAGGCGAAGCTGGGGT..................... 298
        |   ||||||   :::   |||   |||
    192 ySerGluGly...GlnSerSerHisGlyAspGlySerGluPheAspAspG 208

    298 ..... 298

    208 luGlyMetGlnSerAspAspProGluSerThrArgSerAspArgGlyHis 224

    299 .........GCTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGA..... 334
                :::   :::||||||||      |||:::   |||
    225 AlaArgMetSerSerAlaGlyIleArgSerGluGluSerLysGlyAspAr 241

    335 .......ACCCCGACAGAAGATTCTGGACTCCCCAGA............. 364
              |||   |||:::||||||          :::
    241 gGluProThrSerThrGlnAspSerAspAspSerGlnSerValGluPheS 258

    364 ..... 364

    258 erSerArgLysSerPheArgArgSerArgValSerGluGluAspTyrArg 274

    365 .......................CGGGACCAGGAGAGGGACGGCATGAG 390
                               |||:::   :::   |||:::   ::
    275 GlyGluLeuThrAspSerAsnSerArgGluThrGlnSerAspSerThrGl 291

    391 CGACACA...CACAAACACAGAACCACACAGCCAGTCCCAGGAGCCCAGT 437
        :||||||      |||::: |||   :::|||:::||||||||   :::
    291 uAspThrAlaSerLysGlu.GluSerArgSerGluSerGlnGlu...Asp 306

    438 AATGGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTAC 487
        :::::::|||   |||:::::|||   :::   |||::::::|||||::
    307 ThrAlaGluSerGlnSerGlnGluAspSerProGluGlyGlnAspProSe 323

    488 ACCTGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGC 537
        :      ::::::   |||      :::   :::   |||         :
    323 rSerGluSerSerGluGluAlaGlyGluProSerGlnGluSerSerSerG 340
```

```
     538 GCGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGAT 587
         :::::::::||||||      :::|||   :::   :::|||      :::
     340 luSerGlnGluGlyValThrSerGluSerArgGlyAspAsnProAspAsn 356

     588 AAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAAT 626
           |||      |||   ::::::::::::|||::::::
     357 ThrSerGlnThrGlyAspGlnGluAspSerGluSerSer 369

seq_name: SwissProt_39:NR41_HUMAN

seq_documentation_block:
ID   NR41_HUMAN     STANDARD;      PRT;   598 AA.
AC   P22736;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ORPHAN NUCLEAR RECEPTOR HMR (EARLY RESPONSE PROTEIN NAK1) (TR3 ORPHAN
DE   RECEPTOR).
GN   NR4A1 OR HMR OR NAK1 OR GFRP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal skeletal muscle;
RX   MEDLINE=91133413; PubMed=2283997;
RA   Nakai A., Kartha S., Sakurai A., Toback F.G., Degroot L.J.;
RT   "A human early response gene homologous to murine nur77 and rat
RT   NGFI-B, and related to the nuclear receptor superfamily.";
RL   Mol. Endocrinol. 4:1438-1443(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90173205; PubMed=2626032;
RA   Chang C., Kokontis J., Liao S., Chang Y.;
RT   "Isolation and characterization of human TR3 receptor: a member of
RT   steroid receptor superfamily.";
RL   J. Steroid Biochem. 34:391-395(1989).
CC   -!- FUNCTION: ORPHAN NUCLEAR RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: FETAL MUSCLE AND ADULT LIVER, BRAIN, AND
CC       THYROID.
CC   -!- INDUCTION: BY GROWTH-STIMULATING AGENTS.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR4 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D49728; BAA08565.1; -.
DR   EMBL; L13740; AAA36763.1; -.
DR   PIR; A37251; A37251.
DR   HSSP; P19793; 2NLL.
DR   MIM; 139139; -.
DR   InterPro; IPR000536; -.
DR   InterPro; IPR001628; -.
DR   InterPro; IPR001723; -.
DR   InterPro; IPR003070; -.
DR   InterPro; IPR003071; -.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR01284; NUCLEARECPTR.
DR   PRINTS; PR01285; HMRNUCRECPTR.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Phosphorylation.
FT   DNA_BIND     267    332       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      267    287       C4-TYPE.
FT   ZN_FING      303    327       C4-TYPE.
FT   DOMAIN       409    459       LIGAND-BINDING (POTENTIAL).
FT   DOMAIN        82     92       POLY-SER.
FT   DOMAIN       583    586       POLY-PRO.
FT   MOD_RES      341    341       PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT   MOD_RES      351    351       PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT   CONFLICT     262    262       G -> P (IN REF. 2).
FT   CONFLICT     370    370       R -> L (IN REF. 2).
SQ   SEQUENCE   598 AA;  64463 MW;  41DAAEA7C25FDA22 CRC64;


alignment_scores:
            Quality:   99.50            Length:    194
              Ratio:  1.144              Gaps:     10
 Percent Similarity:  44.845   Percent Identity: 28.866

alignment_block:
 US-09-651-563-808/rev x NR41_HUMAN   ..

 Align seg 1/1  to: NR41_HUMAN  from: 1  to: 598

     568 TGCAGCTCTTGCAGATCACCTTCCATGTCGCGCACTGGGATCTCAGCTGT 519
         |||||||||:::   |||   :::   |||   ::::::   ||||||::
      78 CysSerSerAlaSerSerSerAlaSerSerThrSerSerSerSerAlaTh 94

     518 ATCCTGATCTTCTTCTGTC.........................:....T 499
         :|||   :::|||:::                              |
      94 rSerProAlaSerAlaSerPheLysPheGluAspPheGlnValTyrGlyC 111

     498 GCTGCCCAGGT..............GTAGGATCCCGACTTTCAGCTGCT 464
         ||   ||||||               |||   :::   ||||||::::::
     111 ysTyrProGlyProLeuSerGlyProValAspGluAlaLeuSerSerSer 127

     463 GGTTCTTCTTTTTGGGGCTCTCCATTA..................... 437
         ||||||:::::::::|||||||||
     128 GlySerAspTyrTyrGlySerProCysSerAlaProSerProSerThrPr 144

     436 ...................CTGGGCTCCTGGGACTGGCTGTGTGGTTCTG 406
                             |||:::   ||||||         |||
     144 oSerPheGlnProProGlnLeuSerProTrpAspGlySerPheGly.... 159

     405 TGTTTGTGTGTGTCGCTCATGCCGTCCCTCTCCTGGTCCCGTCTGGGGAG 356
                           |||      ||| |||   :::      |||
     160 .................HisPheSerPro.SerGlnThrTyrGluGlyLe 170

     355 TCCAGAATCTTCTGTCGGGGTTCCAGGTGCTTCTGGGACTCGGATACCCC 306
               :::      :::|||   ||||||||        ||||
     170 uArgAlaTrpThrGluGlnLeuProLysAlaSerGly.........ProP 184

     305 CAACAGCACCCCAGCTTCGCCTGAGCCCCTCCACTCCTCGGCCCTCCTTC 256
         || |||   |||:::|||           |||      ||||||
     184 ro.GlnProProAlaPhePheSerPheSerProProThrGlyProSerPr 200

     255 CTCCCTAGTTATGGCCGTGACGAGACAAGAAGGTCCATGGGCGTGGCACC 206
          ||||||:::                 :::::|||
     200 oSerLeuAla.................GlnSerProLeuLysLeuPheP 211

     205 GCATGAGAGGGACGACGACTTCGAGGGCCTTCTCACTCTGAGTCACTGAC 156
           ::::::||| :::      |||   ::::::|||   |||:::
     211 roSerGlnAlaThr.HisGlnLeuGlyGluGlyGluSerTyrSerMetPr 227

     155 CACACAACCCTCCGGAAGCCTGCTGGTTCCTCTTCACTCTCACTCACACT 106
          |||         |||      |||:::|||         |||   :::
     227 oThrAlaPheProGly......LeuAlaProThrSerProHisLeuGluG 242

     105 TCAACTCCCAGTTGGATCTGCCTGTG 80
           ::::::   ||||||   ||||||
     242 lySerGlyIleLeuAspThrProVal 250

seq_name: SwissProt_39:CA2B_HUMAN
```

```
seq_documentation_block:
ID   CA2B_HUMAN     STANDARD;      PRT;  1736 AA.
AC   P13942; Q13273; Q13271; Q13272; Q07751;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 2(XI) CHAIN PRECURSOR.
GN   COL11A2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96032717; PubMed=7559422;
RA   Vuoristo M.M., Pihlajamaa T., Vandenberg P., Prockop D.J.,
RA   Ala-Kokko L.;
RT   "The human COL11A2 gene structure indicates that the gene has not
RT   evolved with the genes for the major fibrillar collagens.";
RL   J. Biol. Chem. 270:22873-22881(1995).
RN   [2]
RP   SEQUENCE OF 59-807 FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=93314796; PubMed=8325374;
RA   Zhidkova N.I., Brewton R.G., Mayne R.;
RT   "Molecular cloning of PARP (proline/arginine-rich protein) from human
RT   cartilage and subsequent demonstration that PARP is a fragment of the
RT   NH2-terminal domain of the collagen alpha 2(XI) chain.";
RL   FEBS Lett. 326:25-28(1993).
RN   [3]
RP   SEQUENCE OF 730-1690 FROM N.A.
RX   MEDLINE=89340485; PubMed=2760050;
RA   Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,
RA   van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;
RT   "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of
RT   cDNA and genomic DNA reveals characteristics of a fibrillar collagen
RT   with differences in genomic organization.";
RL   J. Biol. Chem. 264:13910-13916(1989).
RN   [4]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=95238468; PubMed=7721876;
RA   Zhidkova N.I., Justice S.K., Mayne R.;
RT   "Alternative mRNA processing occurs in the variable region of the
RT   pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";
RL   J. Biol. Chem. 270:9486-9493(1995).
RN   [5]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [6]
RP   VARIANT ARG-661.
RX   MEDLINE=95163096; PubMed=7859284;
RA   Vikkula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,
RA   Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,
RA   van den Hoogen F.H.J., Ropers H.-H., Mayne R., Cheah K.S.E.,
RA   Olsen B.R., Warman M.L., Brunner H.G.;
RT   "Autosomal dominant and recessive osteochondrodysplasias associated
RT   with the COL11A2 locus.";
RL   Cell 80:431-437(1995).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CC       CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
CC   -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
CC       ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
CC       MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD
CC       OF ALPHA 3(XI)=1(II).
CC   -!- ALTERNATIVE PRODUCTS: ALTERNATIVE FORMS OF THIS COLLAGEN ARE
CC       PRODUCED BY ALTERNATIVE SPLICING. THESE FORMS DIFFER IN THE
CC       PRESENCE OR ABSENCE OF THREE SEPARATE REGIONS IN THE N-TERMINAL
CC       NON-COLLAGENOUS DOMAIN.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH
CC       PROTEIN OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING
CC       EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE
CC       CARTILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT
CC       AMOUNTS.
CC   -!- DISEASE: DEFECTS IN COL11A2 ARE THE CAUSE OF AUTOSOMAL DOMINANT
CC       AND RECESSIVE OSTEOCHONDRODYSPLASIAS. (1) THE AUTOSOMAL DOMINANT
CC       FORM OF STICKLER SYNDROME (SS) IS CHARACTERIZED BY MILD
CC       SPONDYLOEPIPHYSEAL DYSPLASIA, OSTEOARTHRITIS, AND SENSORINEURAL
CC       HEARING LOSS. (2) THE AUTOSOMAL RECESSIVE DISORDER IS SIMILAR BUT
CC       MORE SEVERE.
CC   -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32169; AAC50214.1; -.
DR   EMBL; U32169; AAC50213.1; -.
DR   EMBL; U32169; AAC50215.1; -.
DR   EMBL; L18987; AAA35498.1; -.
DR   EMBL; J04974; AAA52034.1; -.
DR   PIR; A32645; A32645.
DR   MIM; 120290; -.
DR   MIM; 108300; -.
DR   MIM; 215150; -.
DR   InterPro; IPR000087; -.
DR   InterPro; IPR000885; -.
DR   Pfam; PF01410; COLFI; 2.
DR   Pfam; PF01391; Collagen; 18.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Alternative splicing; Signal;
KW   Disease mutation.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23   1500       COLLAGEN ALPHA 2(XI) CHAIN.
FT   PROPEP    1501   1736       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN      23    255       NONHELICAL REGION.
FT   DOMAIN     487   1500       TRIPLE-HELICAL REGION.
FT   DOMAIN    1501   1736       NONHELICAL REGION.
FT   DOMAIN      10     18       POLY-LEU.
FT   DOMAIN     298    301       POLY-GLU.
FT   CARBOHYD  1604   1604       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   267    292       MISSING (IN ISOFORMS WITHOUT EXON 6).
FT   VARSPLIC   293    313       MISSING (IN ISOFORMS WITHOUT EXON 7).
FT   VARSPLIC   314    373       MISSING (IN ISOFORMS WITHOUT EXON 8).
FT   VARIANT    661    661       G -> R (IN AUTOSOMAL RECESSIVE
FT                               OSTEOCHONDRODYSPLASIA).
FT                               /FTId=VAR_001907.
FT   CONFLICT    85     85       S -> P (IN REF. 2).
FT   CONFLICT    97     97       Q -> R (IN REF. 2).
FT   CONFLICT   542    542       A -> P (IN REF. 2).
FT   CONFLICT   548    549       MP -> TL (IN REF. 2).
FT   CONFLICT   578    579       AQ -> PR (IN REF. 2).
FT   CONFLICT   704    705       NQ -> KP (IN REF. 2).
FT   CONFLICT   720    720       R -> Q (IN REF. 2).
FT   CONFLICT   726    726       D -> N (IN REF. 2).
FT   CONFLICT   843    846       TGPR -> HGST (IN REF. 3).
FT   CONFLICT   882    884       QGP -> SGS (IN REF. 3).
FT   CONFLICT   894    894       L -> P (IN REF. 3).
FT   CONFLICT  1091   1091       D -> V (IN REF. 3).
FT   CONFLICT  1124   1124       A -> R (IN REF. 3).
FT   CONFLICT  1127   1133       EPGARGP -> GAGGLGT (IN REF. 3).
FT   CONFLICT  1552   1552       E -> R (IN REF. 3).
SQ   SEQUENCE   1736 AA;  171795 MW;  612018674E335150 CRC64;
```

```
alignment_scores:
           Quality:   98.00            Length:    183
             Ratio:  1.167              Gaps:     10
 Percent Similarity:  45.902   Percent Identity: 26.230

alignment_block:
 US-09-651-563-808 x CA2B_HUMAN   ..

 Align seg 1/1  to: CA2B_HUMAN  from: 1  to: 1736

     180 CCTCGAAGTCGTCGTCCCTCTCATGCGGTGCCACGCCCATGGACCTTCTT 229
         |||||||||   ||||||                              ||
     176 ProArgSerAlaArgPro.........................ValLe 183

     230 GTCTCGTCACGGCCATAACTAGGGAGGAAGGAGG.........GCCGAGG 270
         |:::   ||||||           |||   |||            |||
     183 uAspThrHisGlyValIleIlePheGlyAlaArgIleLeuAspGluGluV 200

     271 AGTGGAGGGGCTCA...GGCGAAGCTGGGGTGCTGTTGGGGGTATCCGAG 317
           ::::::|||:::      |||   :::::::::   |||||||:::
     200 alPheGluGlyAspValGlnGluLeuAlaIleValProGlyValGlnAla 216

     318 TCCCAGAAGCACCTGGAACCCCGACAGAAGATTCTGGACTCCCCAGACGG 367
         :::   :::      |||      ||||||   |||:::         ||
     217 AlaTyrGluSerCysGlu......GlnLysGluLeuGluCysGluGlyGl 231

     368 GACCAGGAGAGGGACGGCATGAGCGACACACACAAACACAGAACCACAC. 416
         |   |||:::                          |||   :::||||||
     231 yGlnArgGlu..............ArgProGlnAsnGlnGlnProHisA 243

     417 ...........AGCCAGTCCCAGGAGCCCAGTAAT...GGAGAGCCCCAA 452
                   |||   :::|||:::|||||||:::      :::||||||
     243 rgAlaGlnArgSerProGlnGlnGlnProSerArgLeuHisArgProGln 259

     453 AAAGAAGAACCAGCAGCTGAAAGT..........................  476
         ::::::||||||   ::::::
     260 AsnGlnGluProGlnSerGlnProThrGluSerLeuTyrTyrAspTyrGl 276

     477 ..........................................CGGG 480
                                                    :::|
     276 uProProTyrTyrAspValMetThrThrGlyThrThrProAspTyrGlnA 293

     481 ATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGATC 530
         ||||||||||||||::::::   |||      :::::::::
     293 spProThrProGlyGluGluGluGluIleLeuGluSerSerLeuLeuPro 309

     531 CCAGTGCGCGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACAC 580
         |||::::::::::   :::   |||||||||                ||
     310 ProLeuGluGluGluGlnThrAspLeuGln.........ValProProTh 323

     581 CGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACC 629
         |:::|||:::      |||      ::::::   ||||||   :::|||
     323 rAlaAspArg......PheGlnAlaGluGluTyrGlyGluGlyGlyThr 337

seq_name: SwissProt_39:CYB_LEITA

seq_documentation_block:
ID   CYB_LEITA      STANDARD;      PRT;   371 AA.
AC   P14548;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME B.
GN   COB OR CYTB OR CYB.
OS   Leishmania tarentolae (Sauroleishmania tarentolae).
OG   Mitochondrion.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5689;
RN   [1]
RP   SEQUENCE FROM N.A., AND RNA EDITING.
RX   MEDLINE=88210465; PubMed=2452696;
RA   Shaw J.M., Feagin J.E., Stuart K.D., Simpson L.;
RT   "Editing of kinetoplastid mitochondrial mRNAs by uridine addition and
RT   deletion generates conserved amino acid sequences and AUG initiation
RT   codons.";
RL   Cell 53:401-411(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85297791; PubMed=2994021;
RA   Simpson A.G., Necklemann N., la Cruz V.F., Muhich M.L., Simpson L.;
RT   "Mapping and 5' end determination of kinetoplast maxicircle gene
RT   transcripts from Leishmania tarentolae.";
RL   Nucleic Acids Res. 13:5977-5993(1985).
RN   [3]
RP   SEQUENCE OF 1-48 FROM N.A.
RX   MEDLINE=88124876; PubMed=2448777;
RA   Feagin J.E., Shaw J.M., Simpson L., Stuart K.;
RT   "Creation of AUG initiation codons by addition of uridines within
RT   cytochrome b transcripts of kinetoplastids.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:539-543(1988).
RN   [4]
RP   SEQUENCE OF 21-371 FROM N.A.
RX   MEDLINE=85079995; PubMed=6096360;
RA   de la Cruz V.F., Neckelmann N., Simpson L.;
RT   "Sequences of six genes and several open reading frames in the
RT   kinetoplast maxicircle DNA of Leishmania tarentolae.";
RL   J. Biol. Chem. 259:15136-15147(1984).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC   -!- CAUTION: THE GENOMIC DERIVED SEQUENCE DIFFERS FROM THAT SHOWN AS
CC       IT IS MODIFIED BY EXTENSIVE RNA EDITING.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07542; AAA31879.1; ALT_SEQ.
DR   EMBL; M10126; -; NOT_ANNOTATED_CDS.
DR   EMBL; M19065; AAA31878.1; -.
DR   PIR; H22848; H22848.
DR   PIR; A28118; A28118.
DR   InterPro; IPR000179; -.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL        82     82       IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL        96     96       IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL       183    183       IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL       197    197       IRON 1 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   371 AA;  44555 MW;  0003BD11538EA75B CRC64;


alignment_scores:
           Quality:   97.00            Length:    152
             Ratio:  1.347              Gaps:     10
 Percent Similarity:  47.368   Percent Identity: 23.684

alignment_block:
 US-09-651-563-808/rev x CYB_LEITA   ..

 Align seg 1/1  to: CYB_LEITA  from: 1  to: 371
```

```
     705 GTTGTTTCAGCTTGTCTTCATTTAAACTTGTGGTTGCTCTTCACCTGCTT 656
         :::   ::::::||||||           ||||||:::|||:::|||||
      45 IleCysGlyValCysLeu.........AlaTrpLeuPhePheSerCysPh 58

     655 C............TGGCATTTTACAGTGTTCCTCTTTAGGTATTATCTTC 618
         |           |||:::|||:::::::|||||||:::   :::
      58 eIleCysThrAsnTrpTyrPheValLeuPheLeuTrpAspPhe....... 72

     617 ACCTTGACGCCGGAACCCAAATCCAGATTTATCCCCGGTGTTTGACTGAT 568
              |||   |||      |||:::            |||:::|||:::
      73 .....AspLeuGlyPheValIleArgSerThrHisIleCysPheThrSer 87

     567 GCAGCTCTTGCAGATCACCTTC.......................... 546
           |||:::   :::   :::
      88 Leu.LeuPhePheLeuLeuTyrValHisIlePheLysCysIleValLeuI 104

     545 .................CATGTCGCGCACTGGGATCTCAGCTGTATCCTG 513
                         |||:::      |||   ::::::   |||:::
     104 leIleLeuPheAspThrHisIleLeuValTrpAlaValGlyPheIleIle 120

     512 ATCTTCTTCTGTCTGCTGCCCAGGT...GTAGGATCCCGACTTTCAGCTG 466
         :::||| :::         |||   :::|||      |||   ::::
     121 TyrIlePhe.IleValValIleGlyPheIleGlyTyrValLeuProCysT 137

     465 CTGGTTCTTCTTTTTGGGGC........................... 446
         ::      |||:::||||||
     137 hrMetMetSerTyrTrpGlyLeuThrValPheSerAsnIleLeuAlaThr 153

     445 TCTCCATTACTGGGCTCCTGG......GACTGGCTGTGTGGTTCTGTGTT 402
            |||:::::::|||:::|||         |||:::   ||||||   ::
     154 ValProValIleGlyThrTrpLeuCysTyrTrpIleTrpGlySerGluTy 170

     401 TGTG..............TGTGTCGCTCATGCCGTCCCTCTCCTGGTCC 367
         ::::                        |||::::::   :::|||:::|
     170 rIleAsnAspPheThrLeuLeuLysLeuHisValLeuHisValLeuLeuP 187

     366 CG 365
         ||
     187 ro 187
```

seq_name: SwissProt_39:HMPH_CHICK

```
seq_documentation_block:
ID   HMPH_CHICK     STANDARD;      PRT;   277 AA.
AC   Q05502;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HOMEOBOX PROTEIN PRH.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93087175; PubMed=1360645;
RA   Crompton M.R., Bartlett T.J., Macgregor A.D., Manfioletti G.,
RA   Buratti E., Giancotti V., Goodwin G.H.;
RT   "Identification of a novel vertebrate homeobox gene expressed in
RT   haematopoietic cells.";
RL   Nucleic Acids Res. 20:5661-5667(1992).
CC   -!- FUNCTION: RECOGNIZES THE DNA SEQUENCE 5'-ATTAA-3'. MAY PLAY A ROLE
CC       IN HEMATOPOIETIC DIFFERENTIATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- TISSUE SPECIFICITY: IN ALL HEMATOPOIETIC TISSUES EXCEPT
CC       PERIPHERAL BLOOD ERYTHROCYTES AND IN THE LIVER AND LUNG.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64711; CAA45966.1; -.
DR   HSSP; P22808; 1VND.
DR   TRANSFAC; T02091; -.
DR   InterPro; IPR001356; -.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DOMAIN        1    140       PRO-RICH.
FT   DNA_BIND    144    203       HOMEOBOX.
SQ   SEQUENCE   277 AA;  30213 MW;  BE744C143FB9F9FC CRC64;
```

```
alignment_scores:
            Quality:   96.50            Length:    149
              Ratio:  1.237              Gaps:      8
 Percent Similarity:  52.349   Percent Identity: 26.174

alignment_block:
 US-09-651-563-808 x HMPH_CHICK   ..

 Align seg 1/1  to: HMPH_CHICK  from: 1  to: 277

     210 CCACGCCCATGGACC...TTCTTGTCTCGTCACGGCCATAACTAGGGAGG 256
         |||      |||:::   |||:::::|||      |||:::
     130 ProLeuLeuTrpSerProPheIleGlnArgProLeuHisLys........ 143

     257 AAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGG 306
          |||::::::   :::   |||   |||:::|||      :::   |||
     144 .ArgLysGlyGlyGlnValArgPheSerAsnGluGlnThrIleGluLeuG 160

     307 GGGTATCC...GAGTCCCAGAAGCACCTGGAACCCCGACAG......... 344
             :::   |||:::||||||:::|||:::|||   :::
     160 luLysLysPheGluThrGlnLysTyrLeuSerProProGluArgLysArg 176

     345 ......AAGATTCTG.............................. 353
               |||:::|||
     177 LeuAlaLysLeuLeuGlnLeuSerGluArgGlnValLysThrTrpPheGl 193

     354 ..............................GACTCCCCAGACGGGA 369
                                      ::::::|||::::::|
     193 nAsnArgArgAlaLysTrpArgArgLeuLysGlnGluAsnProGlnAlaT 210

     370 CCAGGAGAGGGACGGCATGAGCGACACACACAAACACAGAACCACACAGC 419
         ||::::::      |||   :::|||     :::   :::|||:::|||
     210 hrLysLysGluGluAlaGluGlyThrGlyAspHisGlyAspProArgSer 226

     420 CAG...TCCCAGGAGCCCAGTAATGGA.........GAGCCCCAAAAAGA 457
         :::   |||   :::|||::::::|||         |||||||||   ::
     227 GluGlySerProSerProAlaGlyGlyGlyGluAlaGluProGlnAspSe 243

     458 AGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAGAAG 507
         :      ||||||::::::::::||||||            :::   ::::
     243 rProSerAlaAlaSerGlnGluAspProGluSerAspValSerAspAspS 260

     508 AAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTGAT 554
         ::||||||::::::              |||:::|||||||||
     260 erAspGlnGluVal.....................AspIleGluGlyAsp 269
```

seq_name: SwissProt_39:YT35_STRFR

```
seq_documentation_block:
ID   YT35_STRFR     STANDARD;      PRT;   348 AA.
AC   P20186;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYPOTHETICAL 35.5 KDA PROTEIN IN TRANSPOSON TN4556.
OS   Streptomyces fradiae.
```

```
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1906;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=Tn4556;
RX   MEDLINE=90185236; PubMed=2155856;
RA   Siemieniak D.R., Slightom J.L., Chung S.T.;
RT   "Nucleotide sequence of Streptomyces fradiae transposable element
RT   Tn4556: a class-II transposon related to Tn3.";
RL   Gene 86:1-9(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29297; AAA88561.1; -.
DR   PIR; JQ0431; JQ0431.
KW   Hypothetical protein; Transposable element.
SQ   SEQUENCE   348 AA;  35520 MW;  3BDF5D83ABBBB92A CRC64;


alignment_scores:
            Quality:   96.00            Length:    142
              Ratio:  1.524              Gaps:      6
 Percent Similarity:  44.366   Percent Identity: 28.169

alignment_block:
 US-09-651-563-808/rev x YT35_STRFR   ..

 Align seg 1/1  to: YT35_STRFR  from: 1  to: 348

     348 TCTTCTGTCGGGGTTCCAGGTGCTTCTGGGACTCGGATACCC..... 307
         |||   :::|||   |||   |||:::|||:::   :::|||
     220 SerArgAlaGlyCysProThrAlaAlaGlySerLeuLeuProAlaProAr 236

     306 .....CCAACAGCAC 297
                    |||:::||||
     236 gProProAlaSerSerAlaSerSerProGlnAlaAlaAlaProAlaAlaP 253

     296 CCCAGCTT.............CGCCTGAGCCCCTCCACTCCTCGGCCC 262
         ||:::                |||   :::||||||:::|||||||||
     253 roSerAlaThrArgLeuProArgArgThrThrProSerAlaProArgPro 269

     261 TCCTTCCTCCCTAGT.............TATGGCCGTGACGAGACAAG 227
         |||      |||:::             :::|||        ||
     270 SerSerArgProAlaArgProProIleProAlaAlaArgProProProAr 286

     226 AAGGTCCATGGGCGTGGCACCGCATGAGAGGGACGACGACTTCGAGGGCC 177
         ||||:::   |||:::    |||           :::::::::||||||
     286 gArgThrProGlyThrPro.Arg.....ProAlaAlaAlaArgAla 299

     176 TTCTCACTCTGAGTCACTGACCACACAACCCTCCGGAAGCCTGCTGGTTC 127
                                    |||   |||||||||
     300 ..............................ArgAlaProAlaGlyCy 305

     126 CTCTTCACTCTCACTCACACTTCAACTCCCAGTTGGATCTGCCTGTGGAC 77
          |||            |||   :::   |||:::        |||
     305 sSerProAlaArgArgThrProSerAlaProThrAspArgArgCysArgA 322

      76 CTACCCGCCGTGTCTCAGTAGCGGAGAAAGAATCCAGACCTCAGGGACCC 27
           :::|||||||   |||               :::||||||   |||
     322 laAlaArgArgGlySerProArgProProAlaAlaArgProProGlyArg 338

      26 GAGTCGCCGGCTCACAGCTCCGCC 3
         ::::::      :::::::||||||
     339 GlnGlyThrArgArgAspSerAla 346

seq_name: SwissProt_39:GGE4_HUMAN

seq_documentation_block:
ID   GGE4_HUMAN      STANDARD;      PRT;   117 AA.
AC   Q13068;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GAGE-4 PROTEIN (G ANTIGEN 4).
GN   GAGE4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=95378788; PubMed=7544395;
RA   van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA   Boon T.;
RT   "A new family of genes coding for an antigen recognized by autologous
RT   cytolytic T lymphocytes on a human melanoma.";
RL   J. Exp. Med. 182:689-698(1995).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC       NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19145; AAA82747.1; -.
DR   MIM; 604246; -.
KW   Multigene family.
SQ   SEQUENCE   117 AA;  12956 MW;  365ED71B2F9DC7AD CRC64;


alignment_scores:
            Quality:   95.50            Length:     90
              Ratio:  1.769              Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 33.333

alignment_block:
 US-09-651-563-808 x GGE4_HUMAN   ..

 Align seg 1/1  to: GGE4_HUMAN  from: 1  to: 117

     411 CCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGA 460
         |||::::::   |||:::|||   :::   :::   |||::::::   ||
      29 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGl 45

     461 ACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAGAAGAAG 510
         |||||||::::::   :::||||||:::   :::|||:::   |||::::
      45 uProAlaThrGlnArgGlnAspProAlaAlaAlaGlnGluGlyGluAspG 62

     511 ATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGCAA 560
         ::      :::|||            :::      |||:::|||   |||
      62 luGlyAlaSerAlaGlyGlnGlyProLysPro...GluAlaAspSerGln 77

     561 GAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCG 610
         |||      :::   :::||||||   :::         |||
      78 GluGlnGlyHisProGlnThrGlyCysGluCysGluAspGlyProAspGl 94

     611 TCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATGCCAGAAGCA 660
          |||      |||    |||:::   |||:::   |||   ||||||
      94 yGlnGluMetAspPro.ProAsnProGluGluValLysThrProGluGlu 110

     661 GGTGAAGAGCAACCACAA 678
         ||||||:::|||   |||
     111 GlyGluLysGlnSerGln 116
```

```
OM of: US-09-651-563-808 to: SPTREMBL_16:*   out_format : pfs

Date: Jun 28, 2001 11:31 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09651563/runat_28062001_145612_7641/app_query.fasta_1.854
 -DB=SPTREMBL_16 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09651563_@CGN1_1_111 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-651-563-808
Query length: 781
Database: SPTREMBL_16:*
Database sequences: 425026
Database length: 132305027
Search time (sec): 77.940000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
sp_human:Q9HD64         +  550.00  872.52  1.6e-40  146   ! Q9hd64 homo sapiens (human). xa
sp_virus:Q9DUN0         +  106.50  155.34  0.2065    976   ! Q9dun0 kaposi's sarcoma-associa
sp_human:Q9P0W9         +  105.50  157.85  0.2764    528   ! Q9p0w9 homo sapiens (human). gl
sp_bacteria:Q9L8W2      +  104.50  158.91  0.3588    355   ! Q918w2 mycoplasma hyorhinis. 87
sp_bacteria:Q9L8V9      +  104.50  158.39  0.3547    384   ! Q918v9 mycoplasma hyorhinis. 87
sp_rodent:Q9ET15        +  104.50  156.14  0.3380    538   ! Q9et15 mus musculus (mouse). re
sp_bacteria:Q9RV01      +  104.00  158.83  0.4034    319   ! Q9rv01 deinococcus radiodurans.
sp_rodent:Q63134        +  103.00  158.98  0.5133    246   ! Q63134 rattus norvegicus (rat).
sp_invertebrate:Q9U229  +  103.00  152.56  0.4472    643   ! Q9u229 caenorhabditis elegans
sp_virus:Q9DUM3         +  102.50  148.58  0.4624   1036   ! Q9dum3 kaposi's sarcoma-associ
sp_virus:Q98148         +  102.50  147.82  0.4549   1162   ! Q98148 kaposi's sarcoma-associ
sp_human:Q9Y5L9         -  102.50  141.55  0.3976   2971   ! Q9y5l9 homo sapiens (human). t
sp_organelle:Q33572     -  102.00  155.04  0.5983    350   ! Q33572 trypanosoma brucei bruce
sp_plant:Q9SNQ8         +  100.50  145.29  0.6933   1054   ! Q9snq8 oryza sativa (rice). hy
sp_invertebrate:Q27033  +  100.00  149.75  0.8594    480   ! Q27033 theileria parva. membr
sp_human:Q15627         -   99.50  151.56  1.01      325   ! Q15627 homo sapiens (human). tr
sp_human:Q9UEU5         +   99.00  157.65  1.29      116   ! Q9ueu5 homo sapiens (human). ga
sp_invertebrate:O76871  +   99.00  144.98  0.9840    773   ! O76871 drosophila melanogaste
sp_invertebrate:Q9W4V0  +   99.00  144.98  0.9840    773   ! Q9w4v0 drosophila melanogaste
sp_invertebrate:Q9I7X0  +   99.00  144.58  0.9755    821   ! Q9i7x0 drosophila melanogaste
sp_plant:Q41516         +   98.50  148.50  1.20      405   ! Q41516 triticum aestivum (wheat
sp_plant:Q9SYY0         +   98.50  148.50  1.20      405   ! Q9syy0 triticum aestivum (wheat
sp_human:Q13059         +   98.00  149.74  1.38      299   ! Q13059 homo sapiens (human). cy
sp_human:Q99866         +   98.00  145.82  1.27      537   ! Q99866 homo sapiens (human). cc
sp_virus:O40947         +   98.00  141.10  1.15     1089   ! O40947 kaposi's sarcoma-associ
sp_human:O43636         +   98.00  140.21  1.13     1245   ! O43636 homo sapiens (human). r
sp_human:Q9UMG2         +   98.00  140.18  1.13     1251   ! Q9umg2 homo sapiens (human). c
sp_human:O15026         -   98.00  138.26  1.08     1668   ! O15026 homo sapiens (human). k
sp_human:Q9UIP9         +   98.00  137.99  1.07     1736   ! Q9uip9 homo sapiens (human). c
sp_virus:Q69118         +   97.50  149.01  1.53      296   ! Q69118 human herpesvirus 4 (eps
sp_invertebrate:Q9VJ23  +   97.00  147.19  1.66      345   ! Q9vj23 drosophila melanogaste
sp_invertebrate:Q9V6Q3  -   97.00  145.93  1.62      417   ! Q9v6q3 drosophila melanogaste
sp_rodent:O88344        +   97.00  145.31  1.59      457   ! O88344 rattus norvegicus (rat).
sp_human:Q9Y2W1         +   97.00  140.39  1.43      955   ! Q9y2w1 homo sapiens (human). th
sp_plant:Q41553         +   96.50  140.66  1.63      815   ! Q41553 triticum aestivum (wheat
sp_virus:Q69130         +   96.00  146.32  2.07      310   ! Q69130 human herpesvirus 4 (eps
sp_invertebrate:Q9VFW1  +   96.00  142.42  1.90      556   ! Q9vfw1 drosophila melanogaste
sp_invertebrate:Q24042  +   96.00  140.06  1.81      791   ! Q24042 drosophila melanogaste
sp_fungi:Q9P536         +   95.50  140.97  2.08      613   ! Q9p536 neurospora crassa. conse
sp_plant:Q9FGK8         -   95.50  135.64  1.85     1361   ! Q9fgk8 arabidopsis thaliana (m
sp_organelle:Q9TED6     -   95.00  148.66  2.76      172   ! Q9ted6 trypanosoma cruzi cru
sp_organelle:Q9TED5     -   95.00  148.66  2.76      172   ! Q9ted5 trypanosoma cruzi cru
sp_virus:Q9DGT6         -   95.00  131.28  1.90     2321   ! Q9dgt6 turkey herpesvirus.
sp_virus:Q66666         +   94.50  127.87  1.99     3436   ! Q66666 equine herpesvirus t
sp_organelle:Q9TED7     -   94.00  147.07  3.38      172   ! Q9ted7 trypanosoma cruzi cru

seq_name: sp_human:Q9HD64

seq_documentation_block:
ID   Q9HD64      PRELIMINARY;      PRT;   146 AA.
AC   Q9HD64;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   XAGE-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Liu X., Helman L.J., Yeung C., Lee B., Pastan I.;
RT   "XAGE-1, A New Gene That is Frequently Expressed in Ewing's Sarcoma.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF251237; AAG01401.1; -.
SQ   SEQUENCE   146 AA;  16163 MW;  4B4FC1E5C7434F21 CRC64;


alignment_scores:
           Quality:  550.00            Length:    227
             Ratio:  4.583               Gaps:      2
 Percent Similarity:  52.863   Percent Identity: 52.423

alignment_block:
 US-09-651-563-808 x Q9HD64   ..

 Align seg 1/1  to: Q9HD64  from: 1  to: 146

       2 CGGCGGAGCTGTGAGCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTC 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      27 ArgArgSerCysGluProAlaThrArgValProGluValTrpIleLeuSe 43

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         ||||||||||||||||||
      43 rProLeuLeuArgHisGly..............................  49

     102 TTGAAGTGTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGGGTTG 151

      49 ..................................................  49

     152 TGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCCTCTC 201

      49 ..................................................  49

     202 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAG 251

      49 ..................................................  49

     252 GGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCT 301

      49 ..................................................  49

     302 GTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTC 351

      49 ..................................................  49

     352 TGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACA 401
                                                   :::  ||||||||
      50 ..................................Gly..HisThrGl 53

     402 AACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
      53 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 70

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      70 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 86

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      87 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 103

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 120

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     120 lySerGlyValLysValLysIleIleProLysGluGluHisCysLysMet 136

     652 CCAGAAGCAGGTGAAGAGCAACCACAAGTT 681
         |||||||||||||||||||||||||||||
     137 ProGluAlaGlyGluGluGlnProGlnVal 146

seq_name: sp_virus:Q9DUN0

seq_documentation_block:
ID   Q9DUN0      PRELIMINARY;      PRT;   976 AA.
AC   Q9DUN0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ORF73.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PK-1;
RX   MEDLINE=99445611; PubMed=10515805;
RA   Gao S.J., Zhang Y.J., Deng J.H., Rabkin C.S., Flore O., Jenson H.B.;
RT   "Molecular polymorphism of Kaposi's sarcoma-associated herpesvirus
RT   (Human herpesvirus 8) latent nuclear antigen: evidence for a large
RT   repertoire of viral genotypes and dual infection with different viral
RT   genotypes.";
RL   J. Infect. Dis. 180:1466-1476(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PK-1;
RX   MEDLINE=20381179; PubMed=10900044;
RA   Zhang Y.J., Deng J.H., Rabkin C., Gao S.J.;
RT   "Hot-spot variations of Kaposi's sarcoma-associated herpesvirus latent
RT   nuclear antigen and application in genotyping by PCR-RFLP.";
RL   J. Gen. Virol. 81:2049-2058(2000).
DR   EMBL; AF192756; AAG01636.1; -.
SQ   SEQUENCE   976 AA;  112017 MW;  E5781E2A509FF70B CRC64;


alignment_scores:
            Quality:  106.50             Length:    107
              Ratio:  1.746                Gaps:      2
 Percent Similarity:  57.009   Percent Identity: 26.168

alignment_block:
 US-09-651-563-808 x Q9DUN0    ..

 Align seg 1/1  to: Q9DUN0  from: 1  to: 976

     261 AGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGGGGGT 310
         :::   |||:::   :::   :::|||:::::::::      |||:::::
     412 LysLysGluAspGluGluAspGlyGlyAspGlyAsnLysThrLeuSerIl 428

     311 ATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTCTGGACTCCC 360
         :::::::||||||::::::   ||||||:::|||:::

     428 eGlnSerSerGlnGlnGlnGlnGluProGlnGlnGln..... 440

     361 CAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACAAACACAGAA 410
                                                        |||
     441 .....Glu 441

     411 CCACACAGC.........CAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         |||:::::::         |||:::|||||||||::::::   ||||||||
     442 ProGlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProGl 458

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         |::::::||||||      |||   ::::::      |||   ||||||
     458 nGlnGlnGluProLeuGlnGluProGlnGlnGlnGluProGlnGlnGlnG 475

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
               ::::::|||:::      :::         ::::::::::|||
     475 luProGlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluPro 491

     552 GATCTGCAAGAGCTGCATCAG 572
         :::   ||||||   :::|||
     492 GlnGlnGlnGluProGlnGln 498

seq_name: sp_human:Q9P0W9

seq_documentation_block:
ID   Q9P0W9      PRELIMINARY;      PRT;   528 AA.
AC   Q9P0W9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   GLUCOSIDASE II BETA SUBUNIT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Marcil A., Sevigny G., Thomas D.Y.;
RT   "Expression of Enzymatically Active Isoforms of Human Glucosidase II
RT   in Insect Cells.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO EF-HAND FAMILY.
DR   EMBL; AF144075; AAF66686.1; -.
DR   InterPro; IPR000886; -.
DR   InterPro; IPR002048; -.
DR   InterPro; IPR002172; -.
DR   Pfam; PF00036; efhand; 1.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
DR   SMART; SM00192; LDLa; 1.
KW   Calcium-binding.
SQ   SEQUENCE   528 AA;  59425 MW;  8DAD9776037E878E CRC64;


alignment_scores:
            Quality:  105.50             Length:    212
              Ratio:  0.995                Gaps:      10
 Percent Similarity:  50.000   Percent Identity: 24.528

alignment_block:
 US-09-651-563-808 x Q9P0W9    ..

 Align seg 1/1  to: Q9P0W9  from: 1  to: 528

      52 TCCGCTACTGAGACACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAG 101
         |||:::||||||      ::::::   |||:::   |||      |||::
     230 SerValThrGluLeu...GlnThrHisProGluLeuAspThrAspGlyAs 245

     102 TTGAAGTGTGAGTGAGAGTGAA..... 123
         :   ::::::||||||:::|||
     245 pGlyAlaLeuSerGluAlaGluAlaGlnAlaLeuLeuSerGlyAspThrG 262

     124 ..GAGGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGTGACTCAGAGT 171
```

```
          :::   :::::::|||         ||||||:::::: :::|||
    262 lnThrAspAlaThrSerPheTyrAspArgValTrpAlaAla.IleArgAs 278

    172 GAGAAGGCCCTCGAAGTCGTCGTCCCTCTCATGCGGTGCCACGCCCATGG 221
         :::      |||:::                        |||   |||   |
    278 pLysTyrArgSerGlu.....................AlaLeuProThrA 288

    222 ACCTTCTTGTCTCGTCACGGCCATAACTAGGGAGG...AAGGAGGGCCGA 268
        |||||   :::   |||   |||   |||   :::   ||||||   :::
    288 spLeuProAlaProSerAlaProAspLeuThrGluProLysGluGluGln 304

    269 GGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGGGGGTATCCGAGT 318

    304 ............................................... 304

    319 CCCAGAAGCACCTGGAACCCCGACAGAAGATTCTGGACTCCCCAGACGGG 368
         |||   :::|||::::::|||||||||:::::          :::::::
    305 .ProProValProSerSerProThrGluGluGlu......GluGluGluG 319

    369 ACCAGGAGAGGGACGGCATGAGCGACACACACAAACACAGAACCACACAG 418
        :::::|||::::::      ::::::   ::::::::: |||      ||
    319 luGluGluGluGluGluAlaGluGluGluGluGlu.GluGluAspSe 335

    419 CCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAG 468
        |:::::::      |||      :::   ||||||      :::||||||
    335 rGluGluAlaProProProLeuSerProProGlnProAlaSerProAlaG 352

    469 CTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGAT 518
          |||::::::   |||         :::|||||:::          |||
    352 luGluAspLysMetProProTyrAspGluGlnThrGlnAlaPheIleAsp 368

    519 ACA............................GCTGAGATCCCAGTGCG 538
        :::                           ||||||      :::::
    369 AlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGluArgSerLeuLy 385

    539 CGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATA 588
        :|||||||||   ::::::::::::|||:::|||                :
    385 sAspMetGluGluSerIleArgAsnLeuGluGln..............G 397

    589 AATCTGGATTTGGGTTCCGGCGTCAAGGTGAA 620
        ::   :::|||   |||      :::||||||
    397 luIleSerPheAspPheGlyProAsnGlyGlu 407

seq_name: sp_bacteria:Q9L8W2

seq_documentation_block:
ID   Q9L8W2      PRELIMINARY;      PRT;   355 AA.
AC   Q9L8W2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   87-KDA SURFACE LIPOPROTEIN PRECURSOR (FRAGMENT).
GN   VLPA87.
OS   Mycoplasma hyorhinis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2100;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SK76;
RX   MEDLINE=20138163; PubMed=10671459;
RA   Citti C., Watson-McKown R., Droesse M., Wise K.S.;
RT   "Gene Families Encoding Phase- and Size-Variable Surface Lipoproteins
RT   of Mycoplasma hyorhinis.";
RL   J. Bacteriol. 182:1356-1363(2000).
DR   EMBL; AF193878; AAF36546.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
SQ   SEQUENCE   355 AA;  35197 MW;  D3D638067B2E117F CRC64;


alignment_scores:
            Quality:  104.50            Length:    241
              Ratio:  0.780               Gaps:     11
 Percent Similarity:  55.602   Percent Identity: 26.141

alignment_block:
 US-09-651-563-808 x Q9L8W2   ..

 Align seg 1/1  to: Q9L8W2  from: 1  to: 355

      16 GCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTCTCCGCTACTGAGAC 65
         :::|||:::::::::      |||:::::|||   ::::::   :::||
      82 SerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGluAsnTh 98

      66 ACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGTG 115
         |::::::             :::      ||||||:::   ::::::||||
      98 rGlnGlnSerGluAlaProGlyThrAsnThrGlyAsnLysThrThrSerG 115

     116 AGAGTGAAGAGGAA..............CCAGCAGGCTTCCGGAGGGTT 150
         |||||::::::|||                ||||||
     115 luSerAsnSerGluSerSerThrGlySerGlnAlaGlyThrThrThrAsn 131

     151 GTGTGGTCAGTGACTCAGAGTGAGAAGGCCCTC......GAAGTCGTCGT 194
         :::   |||   ::::::||||||:::::::::      |||   :::
     132 ThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGl 148

     195 CCCTCTCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCA 244
               ::::::   :::|||   |||   :::          |||
     148 uAsnThrGlnGlnSerGluAlaProGlyThrLys.........ThrGluA 162

     245 TAACTAGGGAGGAAGGAGGGCCGAGGA......GTGGAGGGGCTCAGGCG 288
           |||::::::::::   :::||| |||      :::|||:::   :::::
     162 snThrGlnGlnSerGluAlaPro.GlyThrLysThrGluAsnThrGlnGl 178

     289 AAGCTGGGGTGCTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCC 338
         :|||                                ||||||||||||||
     178 nSer............................GluAlaProGlyThrA 185

     339 CGACAGAAGAT.........TCTGGACTCCCCAGACGGGACCAGGAGAGG 379
           |||   :::          ||| :::|||   :::|||:::      ||
     185 snThrGlyAsnLysThrThrSer.GluSerAsnSerGlySerSerThrGl 201

     380 GACGGCATGAGCGACACACACAAACACAGAACCACACAGCCAGTCCCAGG 429
         |:::      :::|||   |||||||||          |||:::|||::::
     201 ySerGlnAlaGlyThrThrThrAsnThrGlySerGlySerAsnSerGluS 218

     430 AGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGG 479
         ::      |||:::|||      ::::::          :::|||:::
     218 erGlyMetAsnSerGluLysThrGluAsnThrGlnGlnSerGluAla... 233

     480 GATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGAT 529
               ||||||   :::||||||:::   |||:::   :::|||
     234 .........ProGlyThrLysThrGluAsnThrGlnGln...SerGluAl 246

     530 CCCAGTGCGCGACATGGAAGGTGATCTGCAA...GAGCTGCATCAGTCAA 576
          |||            |||:::      |||   |||         :::|
     246 aProGlyThrLysThrGluAsnThrGlnGlnSerGluAlaLeuGlyThrA 263

     577 ACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAAT 626
         |||||||:::|||:::            :::      ::::::   :::
     263 snThrGlyAsnLysThrThrSerGluSerAsnSerGlySerSerThrGly 279

     627 ACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCAC 676
         :::      ||||||      |||:::   |||      ::::::   ::
     280 SerGlnAlaGlyThrThrThrAsnThrGlySerGlySerAsnSerGluSe 296

     677 AAGTTTAAATGAAGACAAG 695
         :::::::|||::::::|||
     296 rGlyMetAsnSerGluLys 302

seq_name: sp_bacteria:Q9L8V9

seq_documentation_block:
ID   Q9L8V9      PRELIMINARY;      PRT;   384 AA.
AC   Q9L8V9;
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   87-KDA SURFACE LIPOPROTEIN PRECURSOR.
GN   VLPA87.
OS   Mycoplasma hyorhinis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2100;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SK76;
RX   MEDLINE=20138163; PubMed=10671459;
RA   Citti C., Watson-McKown R., Droesse M., Wise K.S.;
RT   "Gene Families Encoding Phase- and Size-Variable Surface Lipoproteins
RT   of Mycoplasma hyorhinis.";
RL   J. Bacteriol. 182:1356-1363(2000).
DR   EMBL; AF193880; AAF36549.1; -.
KW   Signal; Lipoprotein.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30    384       87-KDA SURFACE LIPOPROTEIN.
SQ   SEQUENCE   384 AA;  38226 MW;  04A1F0A98E37B79C CRC64;


alignment_scores:
            Quality:  104.50             Length:    241
              Ratio:  0.780                Gaps:     11
 Percent Similarity:  55.602   Percent Identity: 26.141

alignment_block:
 US-09-651-563-808 x Q9L8V9   ..

 Align seg 1/1  to: Q9L8V9  from: 1  to: 384

      16 GCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTCTCCGCTACTGAGAC 65
         :::|||:::::::::      |||::::::|||   ::::::   :::||
     111 SerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGluAsnTh 127

      66 ACGGCGGGTAGGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGTG 115
         |::::::          ::      ||||||:::   ::::::||||
     127 rGlnGlnSerGluAlaProGlyThrAsnThrGlyAsnLysThrThrSerG 144

     116 AGAGTGAAGAGGAA..............CCAGCAGGCTTCCGGAGGGGTT 150
         |||||::::::|||               ||||||
     144 luSerAsnSerGluSerSerThrGlySerGlnAlaGlyThrThrThrAsn 160

     151 GTGTGGTCAGTGACTCAGAGTGAGAAGGCCCTC......GAAGTCGTCGT 194
         :::   |||   ::::::||||||:::::::::      |||   :::
     161 ThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGl 177

     195 CCCTCTCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCA 244
               ::::::   :::|||   |||   :::         |||
     177 uAsnThrGlnGlnSerGluAlaProGlyThrLys.........ThrGluA 191

     245 TAACTAGGGAGGAAGGAGGGCCGAGGA......GTGGAGGGGCTCAGGCG 288
         |||::::::::   :::||| |||      :::|||:::   :::::
     191 snThrGlnGlnSerGluAlaPro.GlyThrLysThrGluAsnThrGlnGl 207

     289 AAGCTGGGGTGCTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCC 338
         :|||                          ||||||||||||||
     207 nSer............................GluAlaProGlyThrA 214

     339 CGACAGAAGAT.........TCTGGACTCCCCAGACGGGACCAGGAGAGG 379
           |||   :::          ||| :::|||   :::|||:::      ||
     214 snThrGlyAsnLysThrThrSer.GluSerAsnSerGlySerSerThrGl 230

     380 GACGGCATGAGCGACACACACAAACACAGAACCACACAGCCAGTCCCAGG 429
         |:::      :::|||   |||||||||         |||:::|||::::
     230 ySerGlnAlaGlyThrThrThrAsnThrGlySerGlySerAsnSerGluS 247

     430 AGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGG 479
         ::      |||:::|||      ::::::         :::|||:::
     247 erGlyMetAsnSerGluLysThrGluAsnThrGlnGlnSerGluAla... 262

     480 GATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAGAT 529
                 ||||||   :::||||||:::   |||:::   :::|||
     263 .........ProGlyThrLysThrGluAsnThrGlnGln...SerGluAl 275

     530 CCCAGTGCGCGACATGGAAGGTGATCTGCAA...GAGCTGCATCAGTCAA 576
           |||            |||:::      |||   |||         :::|
     275 aProGlyThrLysThrGluAsnThrGlnGlnSerGluAlaLeuGlyThrA 292

     577 ACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAAT 626
         ||||||||:::|||:::             :::      ::::::   :::
     292 snThrGlyAsnLysThrThrSerGluSerAsnSerGlySerSerThrGly 308

     627 ACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCAC 676
         :::      ||||||      |||:::   |||      ::::::   ::
     309 SerGlnAlaGlyThrThrThrAsnThrGlySerGlySerAsnSerGluSe 325

     677 AAGTTTAAATGAAGACAAG 695
         :::::::|||::::::|||
     325 rGlyMetAsnSerGluLys 331

seq_name: sp_rodent:Q9ET15

seq_documentation_block:
ID   Q9ET15      PRELIMINARY;      PRT;   538 AA.
AC   Q9ET15;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   RETINITIS PIGMENTOSA GTPASE REGULATOR (FRAGMENT).
GN   RPGR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Vervoort R., Lennon A., Bird A.C., Tulloch B., Axton R., Miano M.G.,
RA   Meindl A., Meitinger T., Ciccodicola A., Wright A.F.;
RT   "Mutational hot spot within a new RPGR exon in X-linked retinitis
RT   pigmentosa.";
RL   Nat. Genet. 0:0-0(2000).
DR   EMBL; AF286473; AAG00552.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   538 AA;  60333 MW;  549F3E092943FDE0 CRC64;


alignment_scores:
            Quality:  104.50             Length:    228
              Ratio:  0.893                Gaps:     11
 Percent Similarity:  51.316   Percent Identity: 25.877

alignment_block:
 US-09-651-563-808 x Q9ET15   ..

 Align seg 1/1  to: Q9ET15  from: 1  to: 538

     109 GTGAGTGAGAGTGAAGAGGAACCAGCAGGCTTCCGGAGG......GTTGT 152
         :::   |||:::::::|||   |||          :::          |||||
       1 IleProGluGluGlnGluGlyProGluAspSerGluGlyAsnValValVa 17

     153 GTGGTCAGTGACTCAGAGTGAGAAGGCC...CTCGAAGTCGTCGTCCCTC 199
         |   :::|||:::|||:::::::|||      ||||||
      17 lGluGlnValValGlnAlaGlnLysGluAsnLeuGluPheGluGlyAspA 34

     200 TCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACG......... 240
               ::::::   |||::::::||||||   ::::::|||
      34 rgLysGluAlaLysAlaGluAlaProSerAspValIleThrGluLysGlu 50

     241 ......GCCATAACTAGGGAGGAAGGAGGGCCGAGGAGTGG.AGGGGCTC 283
                       ||||||:::||||||   |||:::   |||:::::
      51 ValSerGluSerGluArgGluSerGlyGlyGluArgGluAspArgSerGl 67
```

```
    284 AGGCGAAGCTGGGGTGCTGTTG..........................  305
        :|||::::::      :::
     67 uGlyAspGlyAspGlnIleCysGluLysValSerLeuGluThrGluHisL  84

    306 ..............GGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGA  341
                     |||   :::|||   :::|||      :::   |||
     84 euGlnArgAlaGlnGlyLysGlnGluArgLysLysGlyLysAspLysArg  100

    342 CAGAAG...ATTCTGGACTCCCCAGAC..............GGGACCAG  373
           :::   |||||||||      :::              |||   ::
    101 AlaArgCysIleLeuAspMetLysGluArgGluGluAspLysGlyTrpGl  117

    374 GAGAGGGACGGCATGAGCGACACACACAAACACAGAACCACACAGCCAGT  423
        ::::|||:::      :::         :::   :::      :::|||:
    117 uLysGlySerGluGlyGlyAspLysMetLysArgAspGluGlyAsnGlnG  134

    424 CC..................CAGGAGCCCAGTAATGGAGAGCCCCAAAAA  455
        ::                  :::|||   :::   |||:::   ::::::
    134 luLysArgLysLysGluMetGluGluArgAspAlaGlyAspGluArgSer  150

    456 GAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGACA..  503
        ||||||      :::|||::::::::::|||      |||::::::
    151 GluGluGluGluGlyGluGluGluGluProGluGluGlyGluLysGluGl  167

    504 .......GAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGG  546
              ||||||::::::   |||:::|||         |||:::   |
    167 uGlyGlyGluGluGluGluGlyThrSerGluAspGlnSerArgGluAspG  184

    547 AAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGA  596
        ||||||||   ||||||            :::   |||   ::::::|||
    184 luGlyAspArgGlnGlu............LysGluGlyArgArgGluGly  196

    597 TTTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTA  646
           |||   :::::::::|||::::::::::    ||||||
    197 LysGlyArgGlnGluAspGlyArgGluGlyTrp.LysGlu.........  209

    647 AAATGCCAGAAGCAGGTGAAGAGCAACCACAA  678
                    ||||||||||||   |||
    210 ..............GlyGluGluGlnGluGln  215

seq_name: sp_bacteria:Q9RV01

seq_documentation_block:
ID   Q9RV01      PRELIMINARY;      PRT;   319 AA.
AC   Q9RV01;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 33.1 KDA PROTEIN.
GN   DR1229.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001971; AAF10810.1; -.
DR   TIGR; DR1229; -.
KW   Hypothetical protein.
SQ   SEQUENCE   319 AA;  33082 MW;  7F0229C5DAAA2B61 CRC64;

alignment_scores:
            Quality:  104.00              Length:    159
              Ratio:  1.182                 Gaps:     12
 Percent Similarity:  55.346   Percent Identity: 30.818

alignment_block:
 US-09-651-563-808 x Q9RV01   ..

 Align seg 1/1  to: Q9RV01  from: 1  to: 319

     19 GGCGACTCGGGTCCCTGAGGTCTGGATTCTTTCTCCGCTACTGAGACACG 68
        |||:::|||||||||   :::      |||   ::::::|||::::::||
    159 GlySerSerGlyProSerSerGlyCysSerAlaAsnThrThrArgAlaAr 175

     69 GCGGGTAGGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGTGAGA 118
        |::::::      :::::::::|||   :::|||              :
    175 gHisAlaProAlaArgSerSerProLeuAlaSer..............T 187

    119 GTGAAGAGGAACCAGCAGGCTTCCGGA.....................GG 147
        ::   ::::::|||||| |||||||||                     ::
    187 hrGlyHisSerProAla.AlaSerGlyTrpGlyArgSerProSerArgAs 203

    148 GTTGTGTGGTCAGTGACTCAGAGTGAGAAGGCCCTCGAAGTCGTCGTCCC 197
        :   ||||||             :::   |||||||||   ||||||:::
    203 nTrpCysGly............AlaAlaArgProSerValSerSerGlyA 216

    198 TCTCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAA 247
          :::   :::::: |||   ||| |||:::::::   ::::::::::::
    216 rgAlaThrSerThr.ArgCysTrp.ProAlaAlaSerAlaAsnSerValP 232

    248 CTAGGGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGG 297
          :::      ::::::|||||||||      |||   ::::::   |||
    232 roLysProProSerSerProArgSerValArgAlaGlyThrArgTrpGly 248

    298 TGCTGT...TGG......GGGTATCCG...AGTCCCAGAAGCACC..... 330
        ||||||   |||      |||   |||   :::||||||:::|||
    249 CysCysThrTrpThrProGlyAlaProAlaAlaProArgAlaThrThrTh 265

    331 ................TGGAACCCCGACAGAAGATTCTGGACTCCCCAGA 364
                            :::|||:::||||||   |||         |
    265 rArgAsnSerCysSerHisSerProSerArgArgValTrpArg...LeuT 281

    365 CGGGACCAGGAGAGGGACGGCATGAGCGACACACACAAACACAGAACCAC 414
        ||   ||||||                  :::|||   ::::::
    281 hrThrProGly.....................SerThrProSerArg 289

    415 ACAGCCAGTCCCAGGAGCCCA 435
           ||||||   |||   |||
    290 ProAlaSerAlaArgProPro 296

seq_name: sp_rodent:Q63134

seq_documentation_block:
ID   Q63134      PRELIMINARY;      PRT;   246 AA.
AC   Q63134;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CONTIGUOUS REPEAT POLYPEPTIDE PRECURSOR.
GN   CRP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87165965; PubMed=3558393;
RA   Heinrich G., Habener J.F.;
RT   "Genes encoding proteins with homologous contiguous repeat sequences
RT   are highly expressed in the serous cells of the rat submandibular
RT   gland.";
RL   J. Biol. Chem. 262:5262-5270(1987).
DR   EMBL; M31032; AAA40971.1; -.
```

```
DR   EMBL; M17703; AAA40971.1; JOINED.
KW   Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    246       CONTIGUOUS REPEAT POLYPEPTIDE.
SQ   SEQUENCE   246 AA;  26480 MW;  5BD4B8A0477A92DC CRC64;


alignment_scores:
            Quality:  103.00              Length:    139
              Ratio:  1.304                 Gaps:      7
 Percent Similarity:  56.835   Percent Identity: 28.058

alignment_block:
 US-09-651-563-808 x Q63134    ..

 Align seg 1/1  to: Q63134  from: 1  to: 246

     306 GGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTCTGGA 355
         |||   :::|||:::|||:::::::   ||||||   |||         ::
      74 GlySerGluGluGluGlnGlnGlnGlnGluProThrGlnAlaGluAsnGl 90

     356 CTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACAAACA 405
         ::::|||   :::|||                              :::::::
      90 nGluProProAlaThr.........................SerGlyS 98

     406 CAGAACCACACAGCCAGTCCCAGGAGCCCAGT.........AATGGAGAG 446
         ::|||   :::::::|||:::|||||||||:::         |||   |||
      98 erGluGluGluGlnGlnGlnGlnGluProThrGlnAlaGluAsnGlnGlu 114

     447 CCC............................CAAAAAGAAGAACCAGC 466
         |||                           |||::::::::::|||::
     115 ProProAlaThrSerGlySerGluGluGluGlnGlnGlnGlnGlnProTh 131

     467 A...GCTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAGAAGAAGATC 513
         :   ||||||::::::::::|||         :::   :::|||||||:::|
     131 rGlnAlaGluAsnGlnGluProProAlaThrSerGlySerGluGluGluG 148

     514 AGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGCAAGAG 563
         ||:::      |||             :::   :::::::::   ||||||
     148 lnGlnGlnGlnGlu............SerThrGlnAlaGluAsnGlnGlu 160

     564 CTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCA 613
                     :::|||::::::|||   |||   :::   ||
     161 ..............ProSerAspSerAlaGlyGluGlyGlnGluThrGl 172

     614 AGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGT 663
         |   |||:::::::::    |||:::      :::   ||||||   :::
     172 nProGluGluGlyAsn.ValGluSerProProSerSerProGluAsnSer 188

     664 GAAGAGCAACCACAA 678
         :::||||||||||||
     189 GlnGluGlnProGln 193

seq_name: sp_invertebrate:Q9U229

seq_documentation_block:
ID   Q9U229      PRELIMINARY;      PRT;   643 AA.
AC   Q9U229;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Y56A3A.32 PROTEIN.
GN   Y56A3A.32.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matthews L.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
CC   -!- COFACTOR: FAD (BY SIMILARITY).
CC   -!- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
CC       CLASS-I.
DR   EMBL; AL132860; CAB60511.1; -.
DR   InterPro; IPR001100; -.
DR   InterPro; IPR001327; -.
DR   Pfam; PF00070; pyr_redox; 1.
DR   PRINTS; PR00368; FADPNR.
DR   PRINTS; PR00411; PNDRDTASEI.
KW   FAD; Flavoprotein; Oxidoreductase; Redox-active center.
SQ   SEQUENCE   643 AA;  71257 MW;  EAC5A54980A5F75D CRC64;


alignment_scores:
            Quality:  103.00              Length:    141
              Ratio:  1.272                 Gaps:      8
 Percent Similarity:  57.447   Percent Identity: 26.241

alignment_block:
 US-09-651-563-808 x Q9U229    ..

 Align seg 1/1  to: Q9U229  from: 1  to: 643

     207 GTGCCACGCCCA............TGGACCTTCTTGTCTCGTCACGGCCA 244
         :::|||:::|||            |||||||||:::||||||   :::||
      45 IleProLysProGlySerLeuAspTrpThrPhePheSerArg...SerHi 60

     245 TAACTAGGGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTG 294
         |:::   ::::::::::::   |||   ::::::         |||   |
      60 sThrLysSerAlaHisGluPheGluProTyrLys......ProGluIleG 75

     295 GGGTGCTGTTGGGG............GTATCCGAGTCCCAGAAGCACCTG 332
         ||:::::::::|||            ::::::   :::         :::
      75 lyAlaPheIleGlyAlaValAlaPheIleGlyLeuThrLeuIleAlaVal 91

     333 GAACCCCGACAGAAGATTCTGGACTCCCCAGACGGGACCAGGAGAGGGAC 382
               :::      ::::::   :::   |||:::      :::|||
      92 ValIleLysThrAspValPheLysLysGluAspSerHisGlyHisGlyHi 108

     383 GGCATGAGCGACACACACA......AACACAGAACCACACAGCCAGTCC. 425
          :::   |||   |||:::      :::   |||   |||:::|||:::
     108 sGlyHisAlaLysHisSerLysLysHisGluGluLysHisGluGlnLysH 125

     426 ..CAGGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCA..... 467
         ::|||   :::::::::||||||:::|||:::|||   |||
     125 isGluGluLysGluHisAlaGluProGluLysLysGluGluAlaLysPro 141

     468 .........GCTGAAAGTCGGGATCCTACACCTGGGCAG........... 497
                 ||||||   ::::::|||   |||:::|||
     142 GluLysProAlaGluProLysGluProGluProAlaGlnLysGlnAlaGl 158

     498 ..........CAGACAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGC 537
                    |||:::||||||   |||:::|||   :::        :
     158 uGlnProGluGlnAlaGluGluLysGlnGluThrLysAspAlaGluProL 175

     538 GCGACATGGAAGGTGATCTGCAA 560
         ::::::::      |||   |||
     175 ysGluGlnValAspAspArgGln 182

seq_name: sp_virus:Q9DUM3

seq_documentation_block:
ID   Q9DUM3      PRELIMINARY;      PRT;  1036 AA.
AC   Q9DUM3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   LATENT NUCLEAR ANTIGEN (FRAGMENT).
```

```
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Piolot T., Tramier M., Coppey M., Nicolas J.C., Marechal V.;
RT   "Close but distinct regions of LNA1 are responsible for nuclear
RT   targeting and binding to human mitotic chromosomes.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF305694; AAG27458.1; -.
FT   NON_TER    1036   1036
SQ   SEQUENCE   1036 AA;  119328 MW;  2959EDD2C1C042B8 CRC64;


alignment_scores:
            Quality:  102.50             Length:    143
              Ratio:  1.424                Gaps:      5
 Percent Similarity:  50.350   Percent Identity: 23.077

alignment_block:
 US-09-651-563-808 x Q9DUM3   ..

 Align seg 1/1  to: Q9DUM3  from: 1  to: 1036

     261 AGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGGGGGT 310
         :::   |||:::   :::   :::|||:::::::::      |||:::::
     432 LysLysGluAspGluGluAspGlyGlyAspGlyAsnLysThrLeuSerIl 448

     311 ATCCGAGTCCCAGAAGCACCTGGAACCCCGACAGAAGATTCTGGACTCCC 360
         :::::::||||||::::::   ||||||:::|||:::
     448 eGlnSerSerGlnGlnGlnGlnGluProGlnGlnGln............. 460

     361 CAGACGGGACCAGGAGAGGGACGGCATGAGCGACACACACAAACACAGAA 410
                                                        |||
     461 .............................................Glu 461

     411 CCACACAGC.........CAGTCCCAGGAGCCCAGTAATGGAGAGCCCCA 451
         |||::::::         |||:::||||||||||::::::   ||||||||
     462 ProGlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProGl 478

     452 AAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 501
         |::::::||||||      |||   ::::::      |||   ||||||
     478 nGlnGlnGluProLeuGlnGluProGlnGlnGlnGluProGlnGlnGlnG 495

     502 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGT 551
             ::::::|||:::   :::                   :::
     495 luProGlnGlnGlnGluProGlnGln.................GlnGlu 505

     552 GATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGG 601
            |||||||||   :::||| |||
     506 ProLeuGlnGluProGlnGln.GlnGlu..................... 514

     602 GTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATG 651
                                 |||:::::::::::   ::::::
     515 .........................ProGlnGlnGlnGluProGlnGln 522

     652 CCAGAAGCAGGTGAAGAGCAACCACAA 678
         |||      :::::::::||||||
     523 GlnGluProGlnGlnGlnGluProGln 531

seq_name: sp_virus:Q98148

seq_documentation_block:
ID   Q98148      PRELIMINARY;      PRT;  1162 AA.
AC   Q98148;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ORF73 HOMOLOG.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97048116; PubMed=8892957;
RA   Cesarman E., Nador R.G., Bai F., Bohenzky R.A., Russo J.J.,
RA   Moore P.S., Chang Y., Knowles D.M.;
RT   "Kaposi's sarcoma-associated herpesvirus contains G protein-coupled
RT   receptor and cyclin D homologs which are expressed in Kaposi's sarcoma
RT   and malignant lymphoma.";
RL   J. Virol. 70:8218-8223(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97094384; PubMed=8939871;
RA   Moore P.S., Bashoff C., Weiss R.A., Chang Y.;
RT   "Molecular mimicry of human cytokine and cytokine response pathway
RT   genes by KSHV.";
RL   Science 274:1739-1744(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97121480; PubMed=8962146;
RA   Russo J.J., Bohenzky R.A., Chien M.C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT   "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
RT   (HHV8).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Russo J.J., Bohenzky R.A., Chien M.C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Russo J.J., Bohenzky R.A., Chien M.C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U52064; AAC55944.1; -.
DR   EMBL; U75698; AAC57158.1; -.
DR   InterPro; IPR002017; -.
SQ   SEQUENCE   1162 AA;  135213 MW;  1A72CE01C1CB081C CRC64;


alignment_scores:
            Quality:  102.50             Length:    158
              Ratio:  1.235                Gaps:      5
 Percent Similarity:  52.532   Percent Identity: 22.785

alignment_block:
 US-09-651-563-808 x Q98148   ..

 Align seg 1/1  to: Q98148  from: 1  to: 1162

     261 AGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGGGGGT 310
         :::   |||:::   :::   :::|||:::::::::      |||:::::
     422 LysLysGluAspGluGluAspGlyGlyAspGlyAsnLysThrLeuSerIl 438

     311 ATCCGAGTCCCAGAAGCACCTGGAACCC..................... 338
         :::::::||||||::::::   ||||||
     438 eGlnSerSerGlnGlnGlnGlnGluProGlnGlnGlnGluProGlnGlnG 455

     339 ...........CGACAGAAGATTCTGGACTCCCCAGACGGGACCAGGAGA 377
                      :::|||:::   |||::::::|||:::
     455 lnGluProGlnGlnGlnGluProLeuGlnGluProGln............ 467

     378 GGGACGGCATGAGCGACACACACAAACACAGAACCACACAGCCAGTCC.. 425
                                    :::   ||||||::::::|||:::
     468 ........................GlnGlnGluProGlnGlnGlnGluPr 476

     426 ...................CAGGAGCCCAGTAATGGAGAGCCCCAAAAAG 456
                             |||||||||::::::   |||||||||:::
     476 oGlnGlnGlnGluProLeuGlnGluProGlnGlnGlnGluProGlnGlnG 493

     457 AAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGACAGAA 506
         ::||||||      |||   ::::::      |||   ||||||
```

```
     493 lnGluProLeuGlnGluProGlnGlnGlnGluProGlnGlnGlnGluPro 509

     507 GAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTGATCT 556
         ::::::|||:::      :::         ::::::::::|||   :::
     510 GlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProGlnGl 526

     557 GCAAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGGGTTCC 606
          ||||||   :::||| |||   |||
     526 nGlnGluProGlnGln.GlnGluPro..................... 534

     607 GGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAAATGCCAGA 656
                 :::      |||:::::::::::   ::::::   ||
     535 .........GlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGl 548

     657 AGCAGGTGAAGAGCAACCACAA 678
         |      :::::::::||||||
     548 uProGlnGlnArgGluProGln 555
```

seq_name: sp_human:Q9Y5L9

```
seq_documentation_block:
ID   Q9Y5L9      PRELIMINARY;      PRT;  2971 AA.
AC   Q9Y5L9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   TRANSCRIPTIONAL ACTIVATOR SRCAP.
GN   SRCAP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99278407; PubMed=10347196;
RA   Johnston H., Kneer J., Chackalaparampil I., Yaciuk P., Chrivia J.;
RT   "Identification of a novel SNF2/SWI2 protein family member, SRCAP,
RT   which interacts with CREB-binding protein.";
RL   J. Biol. Chem. 274:16370-16376(1999).
CC   -!- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR   EMBL; AF143946; AAD39760.1; -.
DR   InterPro; IPR000330; -.
DR   InterPro; IPR000637; -.
DR   InterPro; IPR001650; -.
DR   InterPro; IPR002965; -.
DR   Pfam; PF00176; SNF2_N; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF02178; AT_hook; 3.
DR   PRINTS; PR00929; ATHOOK.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase.
SQ   SEQUENCE   2971 AA;  315641 MW;  F9F7EE70304B78A3 CRC64;
```

```
alignment_scores:
            Quality:  102.50            Length:    246
              Ratio:  0.827              Gaps:     16
 Percent Similarity:  50.407   Percent Identity: 28.455
```

```
alignment_block:
 US-09-651-563-808/rev x Q9Y5L9    ..

 Align seg 1/1  to: Q9Y5L9  from: 1  to: 2971

     640 GTTCCTCTTTAGGTATTATCTTCACCTTGACGCCGGAACCCAAATCCAGA 591
         ||||||      :::|||:::||||||           |||||||||
    1291 ValProAlaSerAlaLeuAlaSerProPheProSerAlaProAsnProAl 1307

     590 T.................TTATCCCCGGTGTTTGACTGATGCAGCTCTT 559
                           |||:::|||              :::|||
    1307 aProAlaGlnAlaSerLeuLeuAlaPro..............AlaSerS 1319

     558 GCAGATCACCTTCCATGTCGCGCACTGGGATCTCAGCTGTATCCTGATCT 509
              |||   :::::::::                  ::::::   :::
    1319 erAlaSerGlnAlaLeuAlaThrProLeuAlaProMetAlaAlaProGln 1335

     508 TCTTCTGTCTGCTGCCCAGGTGTAGGATCCCGACTTTCAGCTGCTGGTTC 459
         :::::::::   :::|||:::   :::      |||:::
    1336 ThrAlaIleLeuAlaProSerProAlaProProLeuAlaProLeu..... 1350

     458 TTCTTTTTGGGGCTCTCCATTACTGGGCTCCTGGGACTGGCTGTGTG... 412
                       |||:::|||  ||||||:::   |||:::::::
    1351 ................ProValLeu.AlaProSerProGlyAlaAlaPro 1361

     411 GTTCTGTGTTTGTGTGTGTCGCTCATGCCGTCCCTCTCCTGGTCCCGTCT 362
         ||||||:::          :::   :::|||   :::
    1362 ValLeuAlaSerSerGlnThrProValProValMet.............. 1373

     361 GGGGAGTCCAGAATCTTCTGTCGGGGTTCCAGGTGCTTCTGGGACTCGGA 312
             :::|||:::|||:::   |||:::      ||||||::::::   :
    1374 ....AlaProSerSerThrProGlyThrSerLeuAlaSerAlaSerProV 1389

     311 TACCCCCAACAGCACCCCAGCTTCGCCTGAGCCCCTCCACT......... 271
         ::|||      ||||||         |||:::||||||:::
    1389 alPro......AlaProThrProValLeuAlaProSerSerThrGlnThr 1403

     270 ............CCTCGGCCCTCCTTCCTCCCTAGTTATGGCCGTGACGA 233
                     |||   ||||||   ||||||||   :::      ::
    1404 MetLeuProAlaProValProSerProLeuProSerProAlaSerThrGl 1420

     232 GACAAGAAGGTCCATGGGCGTGGCACCGCATGAGAGGGACGACGACTTCG 183
         :|||         :::::::::||||||       :::   ||| |||
    1420 nThr.........LeuAlaLeuAlaPro.AlaLeuAlaProThr.LeuGl 1433

     182 AGGGCCTTCTCACTCTGAGTCACTG.........ACCACACAACCCTCCG 142
          |||   |||   |||::::::|||         |||   :::|||:::|
    1433 yGlySerSerProSerGlnThrLeuSerLeuGlyThrGlyAsnProGlnG 1450

     141 GAAGCCTGCTGGTTCCTCTTCACTCTCACTC.................... 111
         ||  |||      :::::::::||||||||
    1450 ly..ProPheProThrGlnThrLeuSerLeuThrProAlaSerSerLeuV 1466

     110 .................ACACTTCAACTCCCAGTTGGATCTGCCTGTGGA 78
                         ||||||:::|||      |||         |||
    1466 alProThrProAlaGlnThrLeuSerLeuAlaProGlyProProLeuGly 1482

      77 CCTACCCGCCGTGTCTCAGTAGCGGAG............AAAGAATCCAG 40
         |||||||::   :::|||:::|||                    :::
    1483 ProThrGlnThrLeuSerLeuAlaProAlaProProLeuAlaProAlaSe 1499

      39 ACCTCAGGGACCCGAGTCGCCGGCTCACAGC 9
          |||   ||||||   :::||||||||:::
    1499 rProValGlyPro...AlaProAlaHisThr 1508
```

seq_name: sp_organelle:Q33572

```
seq_documentation_block:
ID   Q33572      PRELIMINARY;      PRT;   350 AA.
AC   Q33572;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   KINETOPLAST APOCYTOCHROME B.
GN   PETB OR PETD.
OS   Trypanosoma brucei brucei.
OG   Mitochondrion.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85086128; PubMed=6513991;
RA   Johnson B.J.B., Hill G.C., Donelson J.E.;
RT   "The maxicircle of Trypanosoma brucei kinetoplast DNA encodes
RT   apocytochrome b.";
RL   Mol. Biochem. Parasitol. 13:135-146(1984).
```

```
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q + 2
CC       FERROCYTOCHROME C.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN (BY SIMILARITY).
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
DR   EMBL; M17998; AAA32115.1; -.
DR   InterPro; IPR000179; -.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
KW   Electron transport; Heme; Mitochondrion; Respiratory chain;
KW   Transmembrane.
SQ   SEQUENCE   350 AA;  41984 MW;  9617E3B205FEE9BE CRC64;


alignment_scores:
            Quality:  102.00              Length:    191
              Ratio:  1.275                 Gaps:     12
 Percent Similarity:  41.885   Percent Identity: 23.560

alignment_block:
 US-09-651-563-808/rev x Q33572   ..

 Align seg 1/1  to: Q33572  from: 1  to: 350

     705 GTTGTTTCAGCTTGTCTTCATTTAAACTTGTGGTTGCTCTTCACCTGCTT 656
         :::   ::::::|||||           |||||||:::|||:::|||||
      24 IleCysGlyValCysLeu.........AlaTrpLeuPhePheSerCysPh 37

     655 C............TGGCATTTTACAGTGTTCCTCTTTAGGTATTATCTTC 618
         |            |||:::|||::::::|||||:::   :::
      37 eIleCysSerAsnTrpTyrPheValLeuPheLeuTrpAspPhe....... 51

     617 ACCTTGACGCCGGAACCCAAATCCAGATTTATCCCCGGTGTTTGACTGAT 568
             |||   |||      |||:::           |||:::|||:::
      52 .....AspLeuGlyPheValIleArgSerValHisIleCysPheThrSer 66

     567 GCAGCTCTTGCAGATCACCTTC.......................... 546
           |||      :::   :::
      67 Phe.LeuTyrLeuLeuLeuTyrIleHisIlePheLysSerIleThrLeuI 83

     545 ................CATGTCGCGCACTGGGATCTCAGCTGTATCCTG 513
                       |||:::      |||   ::::::   ||||||
      83 leIleLeuPheAspThrHisIleLeuValTrpPheIleGlyPheIleLeu 99

     512 ...ATCTTCTTCTGTCTGCTGCCCAGGTGTAGGATCCCGACTTTCAGCTG 466
           :::|||:::   ::::::      :::|||      |||   ::::
     100 PheValPheIleIleIleIle.AlaPheIleGlyTyrValLeuProCysT 116

     465 CTGGTTCTTCTTTTTGGGGC........................... 446
         ::      |||:::||||||
     116 hrMetMetSerTyrTrpGlyLeuThrValPheSerAsnIleIleAlaThr 132

     445 TCTCCATTACTGGGCTCCTGG......GACTGGCTGTGTGGTTCTGTGTT 402
           |||:::||||||   |||         |||:::   ||||||   ||
     133 ValProIleLeuGlyIleTrpLeuCysTyrTrpIleTrpGlySerGluPh 149

     401 TGTG..............TGTGTCGCTCATGCCGTCCCTCTCCTGGTCC 367
         |:::                     |||::::::   :::|||:::|
     149 eIleAsnAspPheThrLeuLeuLysLeuHisValLeuHisValLeuLeuP 166

     366 CG.......................................... 365
         ||
     166 roPheIleLeuLeuIleIleLeuIleLeuHisLeuPheCysLeuHisTyr 182

     364 TCTGGGGAGTCCAGAATCTTCTGT.....................CG 339

                 :::|||      ||||||                    ||
     183 PheMetSerSerAspAlaPheCysAspArgPheAlaPheTyrCysGluAr 199

     338 GGGTTCCAGGTGCTTCTGG 320
         |   |||   |||:::|||
     199 gLeuSerPheCysMetTrp 205

seq_name: sp_plant:Q9SNQ8

seq_documentation_block:
ID   Q9SNQ8      PRELIMINARY;      PRT;  1054 AA.
AC   Q9SNQ8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC   Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
RT   clone:P0535G04.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP000399; BAA83569.1; -.
SQ   SEQUENCE   1054 AA;  113434 MW;  97DD6F2144766422 CRC64;


alignment_scores:
            Quality:  100.50              Length:    209
              Ratio:  1.155                 Gaps:      8
 Percent Similarity:  41.627   Percent Identity: 25.359

alignment_block:
 US-09-651-563-808 x Q9SNQ8   ..

 Align seg 1/1  to: Q9SNQ8  from: 1  to: 1054

     123 AGAGGAACCAGCAGGCTTCCGGAGGGTTGTGTGGTCAGTGACTCAGAGTG 172
         ||||||   :::      |||:::|||      :::|||   |||:::
     321 ArgGlyGluGlnSerThrProArgGlyGlyArgAlaSerGlySerArgAs 337

     173 AGAAGGCCCTCGAAGTCGTCGTCCCTCTCATGCG.............. 206
          :::||||||   |||   ||||||:::   |||
     337 pArgGlyProGlySerSerArgProAlaAspAlaArgGlyLysArgLysG 354

     207 .............GTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGC 242
                          |||   |||              |||   |||
     354 lnGlyGlyThrProProProSerPro............ProArgGlyGly 366

     243 CATAACTAGGGAGGAAGGAGGGCCGAGGAGTGGAGGGGCTCAGGCGAAGC 292
               |||:::   ||||||::::::   |||   :::|||   ||
     367 .........GlyAlaAlaArgAlaSerAsnArgArgProGluGlyAlaAl 380

     293 TGGGGTGCTGTTGGGGGTATCCGAGTCCCAGAAGCACCTGGAACCCCGAC 342
         |   :::
     380 aProThr..................................... 382

     343 AGAAGATTCTGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAGCG 392
                 ::::::|||:::|||   |||:::
     383 ...........SerGlnProGluGlyGluArgLysLysLysArgProArg 395

     393 ACACACACAAACACAGAACCACAC.............AGCCAGTCCCA 427
               :::|||||||||           |||      ::
     396 LysMetGlyGluThrGluProSerArgGlyAsnLeuIleSerProProLy 412

     428 GGAGCCCAGTAATGGAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTC 477
         :         |||      |||::::::|||   |||:::      ||||
     412 sTrpSerPheAsnArgProProArgSerGluIleProSerArgProSerA 429
```

```
     478 GGGATCCTACACCTGGGCAGCAGACAGAAGAAGATCAGGATACAGCTGAG 527
         ||            ||||||:::      ||||||      :::||||||
     429 rgHisSerLysSerGlyGlnSerGluAlaGluAspProAlaAlaAlaGlu 445

     528 ATC.................................CCAGTGCGCGA 541
                                             :::|||::
     446 AlaLeuArgArgGluSerAspArgArgGluAlaAlaAspArgLeuArgGl 462

     542 CATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAACACCGGGGATAAAT 591
         :   |||         ||||||      ::::::
     462 uAlaGluGluThrAlaGlnGluAlaAlaArgAla.............. 473

     592 CTGGATTTGGGTTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACA 641
                       ||||||:::|||:::::::: :::||||||
     474 ................ArgGlnAlaGluGluThrAla.ArgGluGluAl 484

     642 CTGTAAAATGCCAGAAGCAGGTGAA 666
          ::::::      :::|||||||||
     484 aAlaArgAlaArgGlnAlaGlyGlu 492

seq_name: sp_invertebrate:Q27033

seq_documentation_block:
ID   Q27033      PRELIMINARY;      PRT;   480 AA.
AC   Q27033;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MEMBRANE PROTEIN.
OS   Theileria parva.
OC   Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC   Theileria.
OX   NCBI_TaxID=5875;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MUGUGA;
RX   MEDLINE=94088665; PubMed=8264721;
RA   Baylis H.A., Allsopp B.A., Hall R., Carrington M.;
RT   "Characterisation of a glutamine- and proline-rich protein (QP
RT   protein) from Theileria parva.";
RL   Mol. Biochem. Parasitol. 61:171-178(1993).
DR   EMBL; L06323; AAA18800.1; -.
KW   Membrane.
SQ   SEQUENCE   480 AA;  52388 MW;  619B46AE128EA2E3 CRC64;


alignment_scores:
            Quality:  100.00             Length:    142
              Ratio:  1.266                Gaps:     10
 Percent Similarity:  55.634   Percent Identity: 28.169

alignment_block:
 US-09-651-563-808 x Q27033   ..

 Align seg 1/1  to: Q27033  from: 1  to: 480

     341 ACAGAAGATTCTGGACTCCCCAGACGGGACCAGGAGAGGGACGGCATGAG 390
         |||:::::::::|||   |||:::   ||||||   :::|||   :::
      68 ThrSerGlnGlnGlyLysProGlnProAspGlnProGlnAsp...GlnPr 83

     391 CGACACACACAAACAC................................ 406
          |||   |||::::::
      83 oAspGlnHisGlnGlnProThrGlnGlyAspThrSerGlyGlnGlnGlyP 100

     407 ..AGAACCACACAGCCAGTCCCAGGAGCCCAGTAATGGAGAGCCCCAAAA 454
               |||   |||||::: ||||||||||||   |||       |||
     100 roAspThrProGlnProIle.GlnGluProSer...GlyProValGlnPr 115

     455 AGAA.....................GAACCAGCAGCTGAAAGTCGGGATC 483
          :::                    |||||||:::   ::::::::::::|
     115 oAspGlnThrGlyGlnGlyProValGluProValAspGlnGlnGlnGlnP 132

     484 CT............ACACCTGGGCAGCAGACAGAAGAAGATCAGGATACA 521
         ||            |||   |||||||||   ::::::   ||||||
     132 roThrGlnGlyAspThrSerGlyGlnGlnGlyGlnGlnProGlnAspGln 148

     522 GCTGAGATCCCAGTGCGCGACATGGAAGGTGATCTGCAAGAGCTGCATCA 571
                   ||||||:::::::::::::|||:::   ::::::
     149 .........ProValGlnGluGlnAspGlyGlnAspSerGlnGlyThrPr 162

     572 GTCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAG 621
          ::::::|||   |||:::||||||          ::::::   |||   |
     162 oGluGlnThrProAspGlnSerGly.........GlnGlnProGlyProA 176

     622 ATAATACCT..............AAAGAGGAACACTGTAAAATGCCAGA 656
         ||::: |||               :::::::::      :::::::|||::
     176 spThr.ProAspGlnProValTyrGlnGlnGlnProValGlnGlnProSe 192

     657 AGCAGGTGAAGAGCAACCACAA 678
         ::::   ::::::|||||||||
     192 rGlyGlnGlnGlnGlnProGln 199
```